Economics Letters 112 (2011) 293–297

Contents lists available at ScienceDirect

Economics Letters

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

Test of random versus fixed effects with small within variation

Jinyong Hahn [a,*], John Ham [b,1], Hyungsik Roger Moon [c,2]

[a] Department of Economics, 8283 Bunche Hall, Mail Stop: 147703, UCLA, Los Angeles, CA 90095, United States
[b] Department of Economics, University of Maryland, Tydings Hall, Room 3105, College Park, MD 20742, United States
[c] Department of Economics, University of Southern California, 3620 South Vermont Avenue, Kaprielian Hall 300, Los Angeles, CA 90089-0253, United States

ARTICLE INFO

Article history:
Received 16 December 2010
Received in revised form
25 April 2011
Accepted 23 May 2011
Available online 2 June 2011

JEL classification:
C12

Keywords:
Hausman test
Within variation
Bootstrap

ABSTRACT

Comparisons of within and between estimators using the conventional Hausman test may be subject to statistical problems if the within variation is not sufficiently large. Adopting an alternative asymptotic approximation, we propose a modification of Hausman test that is valid whether the within variation is small or large.

1. Introduction

With the advent of many panel data sets, researchers are commonly estimating a textbook panel data model with individual effects

$$y_{it} = \alpha_i + x_{it}'\beta + \varepsilon_{it}.$$

In carrying out the estimation, the primary concern of many researchers is whether α_i can be treated as uncorrelated with x_{it}. As is well known, random effects estimation will produce an efficiency gain over fixed effects estimation if α_i is uncorrelated with x_{it}; however, if this condition does not hold, only fixed effects estimation will produce consistent estimates. Hausman (1978) provided a test of random effects versus fixed effects which in principle resolves the dilemma for researchers. However, if the within variation is small, the fixed effect estimates may not be asymptotically normal, potentially invalidating the basic premise of the Hausman test. This problem often arises in empirical work, and when the within variation is likely to be small, researchers almost always use the random effects specification without using the Hausman test as a diagnostic,[3] perhaps because they are concerned that it may not be appropriate in their case. In this paper, we first show that this intuition is theoretically valid in the sense that it is not appropriate to use the conventional Hausman test when some or all of the explanatory variables have little within-person variation. Next, we provide a valid version of the Hausman test of between versus fixed effects for this case. Finally, we show that a version of the bootstrap in fact provides a valid critical value for this test.

2. Conventional comparison of between and within estimators

We consider a textbook panel data model with fixed effects

$$y_{it} = \alpha_i + x_{it}'\beta + \varepsilon_{it},$$

where the x's are time varying strictly exogenous regressors, i.e., $(x_{i1}, \ldots, x_{iT})$ is independent of $(\varepsilon_{i1}, \ldots, \varepsilon_{iT})$. For simplicity, throughout the paper we assume that ε_{it} are i.i.d. over i and t and the individual effect parameter is specified to be

$$\alpha_i = c + \overline{x}_i'\gamma + u_i,$$

where $\overline{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$ and u_i is independent of $(x_{i1}, \ldots, x_{iT})$ and $(\varepsilon_{i1}, \ldots, \varepsilon_{iT})$. The 'between' and 'within' models are

$$\overline{y}_i = \alpha_i + \overline{x}_i'\beta + \overline{\varepsilon}_i \quad \text{and} \qquad (1)$$
$$\widetilde{y}_{it} = \widetilde{x}_{it}'\beta + \widetilde{\varepsilon}_{it},$$

* Corresponding author. Tel.: +1 310 825 2523; fax: +1 310 825 9528.
E-mail addresses: hahn@econ.ucla.edu (J. Hahn), ham@econ.umd.edu (J. Ham), moonr@usc.edu (H.R. Moon).

1 Tel.: +1 202 380 8806; fax: +1 301 405 3542.

2 Tel.: +1 213 740 2108; fax: +1 213 740 8543.

3 See, e.g., Kearney (2005), Sawangfa (2007) and Ham et al. (forthcoming).

0165-1765/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2011.05.018

where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\widetilde{y}_{it} = y_{it} - \bar{y}_i$, etc. In other words, the between model assumes that

$$H_0 : \gamma = 0. \quad (2)$$

The within estimator is $\widetilde{\beta} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{x}_{it}'\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{y}_{it}\right)$ and the between estimator is $\overline{\beta} = \left(\sum_{i=1}^{N}(\bar{x}_i - \bar{x})\left(\bar{x}_i - \bar{x}\right)'\right)^{-1}\left(\sum_{i=1}^{N}(\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})\right)$, where $\bar{x} = \frac{1}{N}\sum_{i=1}^{N}\bar{x}_i$ and $\bar{y} = \frac{1}{N}\sum_{i=1}^{N}\bar{y}_i$. As noted by Hausman (1978), the comparison of random and fixed effect estimators under conventional asymptotics is equivalent to the comparison of the between and within estimators. Letting $\beta_B = \text{plim}\,\overline{\beta} = \beta + \gamma$, it is typically shown that

$$\begin{bmatrix}\sqrt{N}(\widetilde{\beta} - \beta)\\ \sqrt{N}(\overline{\beta} - \beta_B)\end{bmatrix} \Rightarrow \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\Omega_{\widetilde{\beta}} & 0\\ 0 & \Omega_{\overline{\beta}}\end{bmatrix}\right)$$

for some $\Omega_{\widetilde{\beta}}$ and $\Omega_{\overline{\beta}}$. Then the well-known Hausman test statistic is

$$N(\widetilde{\beta} - \overline{\beta})'[\widehat{\Omega}_{\widetilde{\beta}} + \widehat{\Omega}_{\overline{\beta}}]^{-1}(\widetilde{\beta} - \overline{\beta}), \quad (3)$$

where $\widehat{\Omega}_{\widetilde{\beta}}$ and $\widehat{\Omega}_{\overline{\beta}}$ are some consistent estimators of their respective population counterparts. Based on the asymptotic normality of $(\sqrt{N}(\widetilde{\beta} - \beta), \sqrt{N}(\overline{\beta} - \beta_B))$, the asymptotic distribution of the test statistic under the null is understood to be $\chi^2_{\dim(\beta)}$; see Hausman and Taylor (1981).

3. A potential problem with conventional procedure

Implicit in the conventional Hausman test is the assumption that both $\sqrt{N}(\widetilde{\beta} - \beta)$ and $\sqrt{N}(\overline{\beta} - \beta_B)$ are asymptotically normal, which in turn requires that both $\widetilde{x}_{it}$ and $\bar{x}_i$ have sufficient variation. Although the between variation (i.e., $\sum_{i=1}^{N}(\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'$) is typically large, the within variation (i.e., $\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{x}_{it}'$) tends to be small in many applications. As a consequence, the asymptotic normality of $\sqrt{N}(\widetilde{\beta} - \beta)$ may be a dubious assumption, and the conventional test may not be reliable.

To take an extreme example, suppose that $T = 2$ and x_{it} is either zero or one. Write $N = n + m$, and suppose that $x_{i1} < x_{i2}$ for $i = 1, \ldots, m$ and $x_{i1} = x_{i2}$ for $i = m+1, \ldots, n+m$. It is well known that the within estimator can be written as

$$\widetilde{\beta} = \frac{\sum_{i=1}^{N}(x_{i2} - x_{i1})(y_{i2} - y_{i1})}{\sum_{i=1}^{N}(x_{i2} - x_{i1})^2}$$

when $T = 2$. Now because $x_{i2} - x_{i1} = 0$ for $i = n+1, \ldots, n+m$ and $x_{i2} - x_{i1} = 1$ for $i = 1, \ldots, m$, we can see that the within estimator is

$$\widetilde{\beta} = \frac{1}{m}\sum_{i=1}^{m}(y_{i2} - y_{i1}).$$

If m is so small that a sensible asymptotic approximation requires that $n \to \infty$ with m fixed, then the central limit theorem is no longer applicable, and we cannot approximate the within estimator by a normal distribution.[4]

Note that this corresponds to the case where conventional researchers' intuition leads them to forgo the standard Hausman tests if the fixed effect estimates are very noisy.

4. A bootstrap-like solution

Consider the moment condition

$$E\left[\sum_{t=1}^{T}\widetilde{x}_{it}(\widetilde{y}_{it} - \widetilde{x}_{it}'\beta_B)\right] = 0, \quad (4)$$

where $\beta_B = \text{plim}\,\overline{\beta}$ is a solution to the moment equation $E[\bar{x}_i\{\bar{y}_i - E[\bar{y}_i] - (\bar{x}_i - E[\bar{x}_i])'b\}] = 0$. It well known that testing for the null hypothesis (2) in model (1) is equivalent to testing for the moment condition (4).

Define

$$\widehat{\Sigma}_{\widetilde{x}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{x}_{it}', \qquad \widehat{\Sigma}_{\bar{x}} = \frac{1}{N}\sum_{i=1}^{N}(\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})',$$

$$\widehat{\sigma}_{\varepsilon}^2 = \frac{1}{N(T-1)}\sum_{i=1}^{N}\widehat{\widetilde{e}}_i'\widehat{\widetilde{e}}_i, \quad \text{and}$$

$$\widehat{\Omega}_{\overline{\beta}} = \left(\frac{1}{N}\sum_{i=1}^{N}(\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{f}_i^2\right),$$

where $\widehat{\widetilde{e}}_i$ and $\widehat{f}_i$ are defined below in Eqs. (5) and (6). The matrices $\widehat{\Sigma}_{\widetilde{x}}$, and $\widehat{\Omega}_{\overline{\beta}}$ are, respectively, estimates of $\Sigma_{\widetilde{x}}$ and the asymptotic variance of $\sqrt{N}(\overline{\beta} - \beta_B)$, $\Omega_{\overline{\beta}}$, where

$$\Omega_{\overline{\beta}} \equiv \left(\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}(\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'\right)^{-1} \times \left(\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}((\bar{y}_i - \bar{y}) - (\bar{x}_i - \bar{x})'\beta_B)^2\right).$$

A natural statistic for the moment condition (4) is

$$H = \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}(\widetilde{y}_{it} - \widetilde{x}_{it}'\overline{\beta})\right)' \times [\widehat{\sigma}_{\varepsilon}^2\widehat{\Sigma}_{\widetilde{x}} + \widehat{\Sigma}_{\widetilde{x}}\widehat{\Omega}_{\overline{\beta}}\widehat{\Sigma}_{\widetilde{x}}]^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}(\widetilde{y}_{it} - \widetilde{x}_{it}'\overline{\beta})\right).$$

In fact, it is easy to see that when $\widehat{\Omega}_{\widetilde{\beta}} = \widehat{\sigma}_{\varepsilon}^2\widehat{\Sigma}_{\widetilde{x}}^{-1}$, the test statistic H is equivalent to the conventional Hausman test statistic in (3).

In what follows we propose a resampling procedure that approximates the distribution of H even under "weak within variation." (See the next section for rigorous definition on "weak within variation".)

1. Compute
$$\widehat{\widetilde{e}}_i = \widetilde{y}_i - \widetilde{x}_i\widetilde{\beta} - (\widetilde{y} - \widetilde{x}\widetilde{\beta}) \quad (5)$$
$$\widehat{f}_i = \bar{y}_i - \bar{x}_i'\overline{\beta} - (\bar{y} - \bar{x}'\overline{\beta}). \quad (6)$$
(Note that we are de-meaning the residuals.)
2. From the empirical distribution F_N of the sample $\{(\widehat{\widetilde{e}}_i, \widehat{f}_i)\}$, generate a random sample $\{(\widetilde{e}_i^*, f_i^*)\}$.
3. Let
$$H^* = \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{e}_{it}^* - \widehat{\Sigma}_{\widetilde{x}}\widehat{\Sigma}_{\bar{x}}^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\bar{x}_i - \bar{x})f_i^*\right)\right)' \times [\widehat{\sigma}_{\varepsilon}^2\widehat{\Sigma}_{\widetilde{x}} + \widehat{\Sigma}_{\widetilde{x}}\widehat{\Omega}_{\overline{\beta}}\widehat{\Sigma}_{\widetilde{x}}]^{-1}.\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{e}_{it}^* - \widehat{\Sigma}_{\widetilde{x}}\widehat{\Sigma}_{\bar{x}}^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\bar{x}_i - \bar{x})f_i^*\right)\right).$$
4. Repeat 2 and 3 many times and tabulate the distribution of H^*.

[4] We discuss this problem further in our Online Appendix Hahn et al. (2010).

5. Asymptotic theory

We now consider the validity of our procedure under the alternative asymptotics that reflect the small within variation in many applications. We allow for the possibility that components of $\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{x}_{it}'$ may have different rates of convergence. In order to express this idea, we will let δ_k denote the rate of convergence of the (k,k) element of $\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{x}_{it}'$, i.e., we will assume that

$$\frac{1}{N^{\delta_k}}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{k,it}^{2} \tag{7}$$

is stochastically bounded. Write $\Delta_n = \text{diag}(N^{-(1-\delta_1)/2},\dots,N^{-(1-\delta_{\dim(\beta)})/2})$. We note that $\frac{1}{N^{\delta_k}}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{k,it}^{2}$ can be understood to be the (k,k) element of

$$\begin{aligned}\hat{\Lambda} &= \Delta_n^{-1}\widehat{\Sigma}_{\widetilde{x}}\Delta_n^{-1}\\ &= \text{diag}(N^{-\delta_1/2},\dots,N^{-\delta_{\dim(\beta)}/2})\\ &\quad\times\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{x}_{it}'\right)\text{diag}(N^{-\delta_1/2},\dots,N^{-\delta_{\dim(\beta)}/2}).\end{aligned}$$

Below we assume that $\hat{\Lambda}$ has a well-defined limit Λ. Let $\Delta = \lim_n \Delta_n$.

Condition 1 elaborates on the boundedness condition (7), and imposes other regularity conditions:

Condition 1. (a) *ε_{it} are i.i.d. over i and t; (b) $E[\varepsilon_{it}] = 0$ and $E[|\varepsilon_{it}|^8] < \infty$; (c) u_i is independent of $(\varepsilon_{i1},\dots,\varepsilon_{iT})$; (d) $E[u_i] = 0$ and $E[|u_i|^8] < \infty$; (e) $(x_{i1},\dots,x_{iT})$ is a nonstochastic triangular array; (f) $\frac{1}{N}\sum_i \bar{x}_i\bar{x}_i'$ and $\frac{1}{N}\sum_i(\bar{x}_i-\bar{x})(\bar{x}_i-\bar{x})'$ converge to positive definite limits; (g) $\limsup \frac{1}{N}\sum_i\|\bar{x}_i\|^8 < \infty$; (h) $0 \le \delta_k \le 1$ for all $k = 1,\dots,\dim(\beta)$, and $\lim\hat{\Lambda} = \Lambda$, where Λ is a positive definite matrix; (i) $\sigma_\varepsilon^2\Lambda + \Lambda\Delta\Omega_{\overline{\beta}}\Delta\Lambda$ is invertible.*

Let L_N^0 denote the distribution of H under the null. Also, let L_N^* denote the (conditional) distribution of H^* (conditioning on the samples). Then, we show that the distribution of H^* is close to the distribution of H under the null hypothesis (2). This is done by showing that $\rho(L_N^0, L_N^*)$ converges to zero in probability, where $\rho(\cdot,\cdot)$ denotes the Prohorov metric.[5]

Theorem 1. *Under Condition 1, $\rho(L_N^0, L_N^*) = o_p(1)$.*

Proof. In Appendix. □

Theorem 1 establishes that our bootstrap based procedure L_* approximates the L_0 asymptotically, whether the null is correct or not. The approximation does not require that the within variation is large, as is typically assumed in conventional asymptotics. Theorem 1 is valid even when the within variation is so small that $\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{k,it}^{2}$ is fixed as $N\to\infty$. (Condition 1(h) allows the possibility that $\delta_k = 0$.) On the other hand, Theorem 1 is valid when the within variation is large. (Condition 1(h) also allows the possibility that $\delta_k = 1$.) Therefore, our bootstrap-like procedure is robust to the degree of within variation, unlike the conventional comparison of the within and between estimators discussed in Section 3.

[5] The definition of the Prohorov metric is provided in the Online Appendix of Hahn et al. (2010). The Prohorov metric metrizes weak convergence on any separable metric space such as Euclidean space. See, e.g., Dudley (1989, Theorem 11.3.3).

Appendix. Proof of Theorem 1

Let $\Gamma_p = \Gamma_p(B)$ be the set of probabilities ν on a Borel σ-field of B such that $\int\|z\|^p\nu(dz) < \infty$. For $\nu,\nu^*\in\Gamma_p$, let $d_p(\nu,\nu^*)$ be the infimum of $E\{\|Z-Z^*\|^p\}^{1/p}$ over pairs of B-valued random variables Z and Z^*, such that $Z\sim\nu$ and $Z^*\sim\nu^*$. By Lemma 8.1 of Bickel and Freedman (1981), the infimum is attained and d_p is a metric on Γ_p. Write $\widetilde{x}_i = (\widetilde{x}_{i1},\dots,\widetilde{x}_{iT})'$ and $\widetilde{\varepsilon}_i = (\widetilde{\varepsilon}_{i1},\dots,\widetilde{\varepsilon}_{iT})'$.

We begin with a few lemmas. Proofs of Lemmas 1–3 are available in our Online Appendix (Hahn et al., 2010).

Lemma 1. *Under Condition 1, (a) $(\widetilde{\beta}-\beta)'\left(\frac{1}{N}\sum_{i=1}^{N}\widetilde{x}_i'\widetilde{x}_i\right)(\widetilde{\beta}-\beta) = o_{a.s}(1)$, and (b) $(\overline{\beta}-\beta_B)'\left(\frac{1}{N}\sum_{i=1}^{N}(\bar{x}_i-\bar{x})(\bar{x}_i-\bar{x})'\right)(\overline{\beta}-\beta_B) = o_{a.s}(1)$.*

Lemma 2. *Under Condition 1, $\widehat{\sigma}_\varepsilon^2 = \frac{1}{N(T-1)}\sum_{i=1}^{N}\widehat{e}_i'\widehat{e}_i = \sigma_\varepsilon^2 + o_{a.s}(1)$.*

Lemma 3. *Under Condition 1, $\widehat{\Omega}_{\overline{\beta}} = \Omega_{\overline{\beta}} + o_{a.s.}(1)$.*

Lemma 4. *Let $\psi_1 \equiv \Delta_n^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{\varepsilon}_{it}\right)$, $\psi_1^* \equiv \Delta_n^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{e}_{it}^*\right)$, $\psi_2 \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\bar{x}_i-\bar{x})\zeta_i$, and $\psi_2^* \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\bar{x}_i-\bar{x})f_i^*$, where $\zeta_i \equiv u_i+\bar{\varepsilon}_i$. Also let $\psi = (\psi_1',\psi_2')'$ and likewise for ψ^*. Suppose that Condition 1 holds. Then, $d_2(\psi,\psi^*) = o_{a.s.}(1)$.*

Proof. Notice that

$$\begin{aligned} d_2(\psi,\psi^*)^2 &= d_2\left(\begin{bmatrix}\Delta_n^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{x}_i'\widetilde{\varepsilon}_i\\ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\bar{x}_i-\bar{x})\zeta_i\end{bmatrix},\right.\\ &\qquad\left.\begin{bmatrix}\Delta_n^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{x}_i'\widetilde{e}_i^*\\ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\bar{x}_i-\bar{x})f_i^*\end{bmatrix}\right)^2\\ &\le \sum_{i=1}^{N} d_2\left(\begin{bmatrix}\Delta_n^{-1}\frac{1}{\sqrt{N}}\widetilde{x}_i'\widetilde{\varepsilon}_i\\ \frac{1}{\sqrt{N}}(\bar{x}_i-\bar{x})\zeta_i\end{bmatrix},\right.\\ &\qquad\left.\begin{bmatrix}\Delta_n^{-1}\frac{1}{\sqrt{N}}\widetilde{x}_i'\widetilde{e}_i^*\\ \frac{1}{\sqrt{N}}(\bar{x}_i-\bar{x})f_i^*\end{bmatrix}\right)^2\end{aligned}$$

by Bickel and Freedman (1981, Lemma 8.7). Denote $\xi_i = (\widetilde{\varepsilon}_i', \zeta_i')'$ and $\widehat{\xi}_i^* = (\widetilde{e}_i^{*\prime}, f_i^{*\prime})'$. Use Eqs. (8.2) and (8.3) of Bickel and Freedman (1981) and bound the RHS by

$$\begin{aligned}&\sum_{i=1}^{N}\left(\left\|\Delta_n^{-1}\frac{1}{\sqrt{N}}\widetilde{x}_i\right\|^2 + \left\|\frac{1}{\sqrt{N}}(\bar{x}_i-\bar{x})\right\|^2\right)d_2\left(\xi_i,\widehat{\xi}_i^*\right)^2\\ &= \left(\text{trace}\left(\Delta_n^{-1}\frac{1}{N}\sum_{i=1}^{N}\widetilde{x}_i'\widetilde{x}_i\Delta_n^{-1}\right)\right.\\ &\quad\left.+\,\text{trace}\left(\frac{1}{N}\sum_{i=1}^{N}(\bar{x}_i-\bar{x})(\bar{x}_i-\bar{x})'\right)\right)\cdot d_2(\xi_i,\widehat{\xi}_i^*)^2\\ &= O(1)d_2(\xi_i,\widehat{\xi}_i^*)^2,\end{aligned}$$

where the first equality holds since $(\xi_i, \widehat{\xi}_i^*)$ are identically distributed and the second equality holds by Condition 1(h). For the required result, it remains to prove that $d_2(\xi_i, \widehat{\xi}_i^*) = o_{a.s.}(1)$. Let $\{\xi_i^*\}_{i=1,\ldots,N}$ denote the i.i.d. samples from the empirical distribution of $\{\xi_i\}_{i=1,\ldots,N}$. Then, by the triangle inequality, $d_2(\xi_i, \widehat{\xi}_i^*) \le d_2(\xi_i, \xi_i^*) + d_2(\xi_i^*, \widehat{\xi}_i^*)$. By Lemma 8.4 of Bickel and Freedman (1981), $d_2(\xi_i, \xi_i^*) = o_{a.s.}(1)$. Next, we apply Lemma 8.8 of Bickel and Freedman (1981) twice to obtain

$$\begin{aligned} d_2(\xi_i^*, \widehat{\xi}_i^*)^2 &= d_2(\xi_i^* - \overline{\xi}, \widehat{\xi}_i^*)^2 + \|\overline{\xi}\|^2 \\ &= d_2\left(\xi_i^*, \widehat{\xi}_i^* + \begin{bmatrix} \overline{\widetilde{y}} - \overline{\widetilde{x}}\widetilde{\beta} \\ \overline{y} - \overline{x}'\overline{\beta} \end{bmatrix}\right)^2 \\ &\quad - \left\| \overline{\xi} - \begin{bmatrix} \overline{\widetilde{y}} - \overline{\widetilde{x}}\widetilde{\beta} \\ \overline{y} - \overline{x}'\overline{\beta} \end{bmatrix} \right\|^2 + \|\overline{\xi}\|^2 \\ &= d_2\left(\xi_i^*, \widehat{\xi}_i^* + \begin{bmatrix} \overline{\widetilde{y}} - \overline{\widetilde{x}}\widetilde{\beta} \\ \overline{y} - \overline{x}'\overline{\beta} \end{bmatrix}\right)^2 \\ &\quad - \left\| \begin{bmatrix} \overline{\widetilde{\varepsilon}} \\ \overline{\zeta} \end{bmatrix} - \begin{bmatrix} \overline{\widetilde{y}} - \overline{\widetilde{x}}\widetilde{\beta} \\ \overline{y} - \overline{x}'\overline{\beta} \end{bmatrix} \right\|^2 + \|\overline{\xi}\|^2. \end{aligned}$$

By definition

$$\begin{aligned} d_2&\left(\xi_i^*, \widehat{\xi}_i^* + \begin{bmatrix} \overline{\widetilde{y}} - \overline{\widetilde{x}}\widetilde{\beta} \\ \overline{y} - \overline{x}'\overline{\beta} \end{bmatrix}\right)^2 \le \frac{1}{N}\sum_{i=1}^N \left\| \begin{bmatrix} \widetilde{\varepsilon}_i - \widetilde{y}_i + \widetilde{x}_i\widetilde{\beta} \\ \zeta_i - \overline{y}_i + \overline{x}_i'\overline{\beta} \end{bmatrix} \right\|^2 \\ &= (\widetilde{\beta} - \beta)'\left(\frac{1}{N}\sum_{i=1}^N \widetilde{x}_i'\widetilde{x}_i\right)(\widetilde{\beta} - \beta) + (\overline{\beta} - \beta_B)' \\ &\quad \times \left(\frac{1}{N}\sum_{i=1}^N \overline{x}_i\overline{x}_i'\right)(\overline{\beta} - \beta_B) = o_{a.s.}(1), \end{aligned}$$

where the last equality is based on Lemma 1. Also,

$$\begin{aligned} \|\overline{\widetilde{\varepsilon}} - \overline{\widetilde{y}} + \overline{\widetilde{x}}'\widetilde{\beta}\|^2 &\le \frac{1}{N}\sum_{i=1}^N \|\widetilde{\varepsilon}_i - \widetilde{y}_i + \widetilde{x}_i\widetilde{\beta}\|^2 \\ &\quad \text{(by the Cauchy–Schwarz inequality)} \\ &= \frac{1}{N}\sum_{i=1}^N \|\widetilde{x}_i(\widetilde{\beta} - \beta)\|^2 = o_{a.s.}(1), \end{aligned}$$

and likewise, $\|\overline{\zeta} - \overline{y} + \overline{x}'\overline{\beta}\| = o_{a.s.}(1)$. Finally, $\|\overline{\xi}\|^2 = \left\|\frac{1}{N}\sum_{i=1}^N \xi_i\right\|^2 = o_{a.s.}(1)$ by SLLN. Therefore, we deduce $d_2(\xi_i^*, \widehat{\xi}_i^*) = o_{a.s.}(1)$, and $d_2(\psi, \psi^*) = o_{a.s.}(1)$, as required. □

Proof of Theorem 1. The 'denominator' of H can be rewritten as $\widehat{\sigma}_\varepsilon^2\widehat{\Sigma}_{\widetilde{x}} + \widehat{\Sigma}_{\widetilde{x}}\widehat{\Omega}_{\overline{\beta}}\widehat{\Sigma}_{\widetilde{x}} = \Delta_n(\widehat{\sigma}_\varepsilon^2\widehat{\Lambda} + \widehat{\Lambda}\Delta_n\widehat{\Omega}_{\overline{\beta}}\Delta_n\widehat{\Lambda})\Delta_n$. By Lemmas 2 and 3, and Condition 1(h), we have

$$\widehat{\sigma}_\varepsilon^2\widehat{\Lambda} + \widehat{\Lambda}\Delta_n\widehat{\Omega}_{\overline{\beta}}\Delta_n\widehat{\Lambda} = \sigma_\varepsilon^2\Lambda + \Lambda\Delta\Omega_{\overline{\beta}}\Delta\Lambda + o_{a.s.}(1), \tag{8}$$

where $\Delta \equiv \lim \Delta_n$. By definition $\widehat{\Sigma}_{\widetilde{x}} = \frac{1}{N}\sum_{i=1}^N\sum_{t=1}^T \widetilde{x}_{it}\widetilde{x}_{it}' = \Delta_n\widehat{\Lambda}\Delta_n$. Let $H_{\text{null}} = \Psi'(\widehat{\sigma}_\varepsilon^2\widehat{\Lambda} + \widehat{\Lambda}\Delta_n\widehat{\Omega}_{\overline{\beta}}\Delta_n\widehat{\Lambda})^{-1}\Psi$, where

$$\begin{aligned} \Psi &\equiv \Delta_n^{-1}\left(\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N\sum_{t=1}^T \widetilde{x}_{it}\widetilde{\varepsilon}_{it}\right) - \widehat{\Sigma}_{\widetilde{x}}\widehat{\Sigma}_{\overline{x}}^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \overline{x}_i\zeta_i\right)\right) \\ &= \Delta_n^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N\sum_{t=1}^T \widetilde{x}_{it}\widetilde{\varepsilon}_{it}\right) - \widehat{\Lambda}\Delta_n\widehat{\Sigma}_{\overline{x}}^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \overline{x}_i\zeta_i\right) \\ &= \psi_1 - \widehat{\Lambda}\Delta_n\widehat{\Sigma}_{\overline{x}}^{-1}\psi_2. \end{aligned}$$

Note that the distribution of H_{null} is equal to the null distribution L_N^0 of H. Also, write $H^* = (\Psi^*)'(\widehat{\sigma}_\varepsilon^2\widehat{\Lambda} + \widehat{\Lambda}\Delta_n\widehat{\Omega}_{\overline{\beta}}\Delta_n\widehat{\Lambda})^{-1}\Psi^*$, where

$$\begin{aligned} \Psi^* &\equiv \Delta_n^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N\sum_{t=1}^T \widetilde{x}_{it}\widetilde{e}_{it}^* - \widehat{\Sigma}_{\widetilde{x}}\widehat{\Sigma}_{\overline{x}}^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N(\overline{x}_i - \overline{x})f_i^*\right)\right) \\ &= \Delta_n^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N\sum_{t=1}^T \widetilde{x}_{it}\widetilde{e}_{it}^*\right) \\ &\quad - \widehat{\Lambda}\Delta_n\widehat{\Sigma}_{\overline{x}}^{-1}g\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N(\overline{x}_i - \overline{x})f_i^*\right) \\ &= \psi_1^* - \widehat{\Lambda}\Delta_n\widehat{\Sigma}_{\overline{x}}^{-1}\psi_2^*. \end{aligned}$$

Note that the (conditional) distribution of H^* (conditioning on the sample observations) is equal to L_N^*.

Denote Z^N to be the samples. Denote $P^*(\cdot|Z^N)$ to be the bootstrap distribution of H^* given samples Z^N (this is the law L_N^*). For every continuity point t, we will show that

$$P^*(H^* \le t|Z^N) = P(H_{\text{null}} \le t) + o_p(1). \tag{9}$$

Define $\rho(P, Q) = \inf\{\varepsilon > 0 : P(A) \le Q(A^\varepsilon) + \varepsilon \text{ for all Borel sets } A\}$, where $Q(A^\varepsilon)$ denotes the ε-inflated set of A (see page 309 of Dudley (1989)). For (9), it suffices to show[6] that $\rho(L_N^0, L_N^*) = o_p(1)$, for which it is enough to show that for every subsequence $\rho(L_N^0(n'), L_N^*(n'))$, there exists a further subsequence $\rho(L_N^0(n''), L_N^*(n''))$ such that $\rho(L_N^0(n''), L_N^*(n'')) = o_{a.s.}(1)$. (See Theorem 9.2.1 of Dudley (1989).)

Notice that $\psi = (\psi_1, \psi_2) = O_p(1)$. Then, given any subsequence n', we can find a further subsequence n'' such that $\psi(n'') = (\psi_1(n''), \psi_2(n'')) \Rightarrow \psi = (\psi_1, \psi_2)$, where the limit may depend on n'. We denote this limit by $\psi(n')$. By Lemma 4, we have

$$d_2(\psi(n''), \psi^*(n'')) = o_{a.s.}(1) \tag{10}$$

along the subsubsequence n''.

Also, notice by definition that

$$\psi(n'') \Rightarrow \psi(n'). \tag{11}$$

Furthermore, under Condition 1 it is easy to show that $\{\|\psi(n'')\|^2\}$ is uniformly integrable, which yields together with (11) the second moment convergence. Then, by Lemma 8.3 of Bickel and Freedman (1981), we have

$$d_2(\psi(n''), \psi(n')) = o(1). \tag{12}$$

Since $d_2(\psi^*(n''), \psi(n')) \le d_2(\psi(n''), \psi^*(n'')) + d_2(\psi(n''), \psi(n'))$, from (10) and (12), we deduce $d_2(\psi^*(n''), \psi(n')) = o_{a.s.}(1)$. Then, by Lemma 8.3 of Bickel and Freedman (1981) we have $\psi^*(n'') \Rightarrow \psi(n')$. Denote $\Psi(n') = \psi_1(n') - \Lambda\Delta_n\Sigma_{\overline{x}}^{-1}\psi_2(n')$. We now use (8) and the continuous mapping theorem to deduce that

$$\begin{aligned} H_{\text{null}} &= \Psi'(\widehat{\sigma}_\varepsilon^2\widehat{\Lambda} + \widehat{\Lambda}\Delta_n\widehat{\Omega}_{\overline{\beta}}\Delta_n\widehat{\Lambda})^{-1}\Psi \\ &\Rightarrow \Psi(n')'(\sigma_\varepsilon^2\Lambda + \Lambda\Delta\Omega_{\overline{\beta}}\Delta\Lambda)^{-1}\Psi(n') = \mathcal{L}(n'), \quad \text{say} \end{aligned}$$

along the subsubsequence n''. Similarly, we can show that $H^* \Rightarrow \mathcal{L}(n')$ a.s. along the subsubsequence n''. Then, by Theorem 11.3.3. of Dudley (1989), since $H_{\text{null}} \Rightarrow \mathcal{L}(n')$ implies $\rho(H_{\text{null}}, \mathcal{L}(n')) = o(1)$ and $H^* \Rightarrow \mathcal{L}(n')$ a.s. implies $\rho(H^*, \mathcal{L}(n')) = o_{a.s.}(1)$ along the subsubsequence n'' and we can claim that

$$\begin{aligned} \rho(L_N^0(n''), L_N^*(n'')) &\le \rho(L_N^0(n''), \mathcal{L}(n')) + \rho(\mathcal{L}(n'), L_N^*(n'')) \\ &= o_{a.s}(1), \end{aligned}$$

as required. □

[6] We can show the result by modifying the proof of Theorem 11.3.3 of Dudley (1989), Part $(d) \Rightarrow (a)$.

References

Bickel, P.J., Freedman, D.A., 1981. Some asymptotic theory for the bootstrap. Annals of Statistics 9, 1196–1217.

Dudley, R.M., 1989. Real Analysis and Probability. Pacific Grove, Wadsworth & Brooks/Cole.

Hahn, J., Ham, J.C., Moon, H.R., 2010. Online appendix. Located at: www-rcf.usc.edu/~moonr.

Ham, J.C., Li, X., Shore-Sheppard, L.S., 2009. A re-examination of the impact of welfare reform on health insurance among less-skilled women. In: James P. Ziliak (Ed.), Welfare Reform and its Long-Term Consequences for America's Poor, Cambridge, UK: Cambridge University Press (forthcoming).

Hausman, J.A., 1978. Specification tests in econometrics. Econometrica 46, 1273–1291.

Hausman, J.A., Taylor, W.E., 1981. Panel data and unobservable individual effects. Econometrica 49, 1377–1398.

Kearney, M.S., 2005. State lotteries and consumer behavior. Journal of Public Economics 89, 2269–2299.

Sawangfa, O., 2007. Job-switching and job satisfaction in South Korea. Mimeo, USC Department of Economics.