Title: Test of Random vs Fixed Effects with Small Within Variation

Author 1 (Corresponding Author)
Name: Jinyong Hahn
Affiliation: UCLA
Address: Department of Economics, 8283 Bunche Hall, Mail Stop: 147703, UCLA, Los Angeles, CA 90095
e-mail: hahn@econ.ucla.edu
Telephone: (310) 825-2523
Fax: (310) 825-9528

Author 2
Name: John Ham
Affiliation: University of Maryland
Address: Department of Economics, University of Maryland, Tydings Hall, Room 3105, College Park, MD 20742,
e-mail: ham@econ.umd.edu
Telephone: (202) 380-8806
Fax: (301) 405-3542

Author 3
Name: Hyungsik Roger Moon
Affiliation: University of Southern California
Address: Department of Economics, University of Southern California, 3620 South Vermont Avenue, Kaprielian Hall 300, Los Angeles, CA 90089-0253
e-mail: moonr@usc.edu
Telephone: (213) 740-2108
Fax: (213) 740-8543

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.
Test of Random vs Fixed Effects with Small Within Variation

Jinyong Hahn  John Ham  Hyungsik Roger Moon
UCLA  University of Maryland  University of Southern California

December 13, 2010

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

\[ y_{it} = \alpha_i + x_{it}' \beta + \epsilon_{it}. \]

In carrying out the estimation, the primary concern of many researchers is whether \( \alpha_i \) 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 \( \alpha_i \) 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 effects 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,\(^1\) 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.

\(^1\)See, e.g., Kearney (2005), Sawangfa (2007) and Ham, Li and Shore-Sheppard (2009).
2 Conventional Comparison of Between and Within Estimators

We consider a textbook panel data model with fixed effects

\[ y_{it} = \alpha_i + x_{it}' \beta + \epsilon_{it}, \]

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

\[ \alpha_i = c + \bar{x}_i' \gamma + u_i, \]

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

\[ \bar{y}_i = \alpha_i + \bar{x}_i' \beta + \bar{\epsilon}_i \]

and

\[ \tilde{y}_{it} = \bar{x}_i' \beta + \tilde{\epsilon}_{it}, \]

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

\[ H_0 : \gamma = 0. \]

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

\[ \left[ \begin{array}{c} \sqrt{N} \left( \tilde{\beta} - \beta \right) \\ \sqrt{N} \left( \bar{\beta} - \beta_B \right) \end{array} \right] \Rightarrow \mathcal{N} \left( \left[ \begin{array}{c} 0 \\ 0 \end{array} \right], \left[ \begin{array}{cc} \Omega_{\tilde{\beta}} & 0 \\ 0 & \Omega_{\bar{\beta}} \end{array} \right] \right) \]

for some \( \Omega_{\tilde{\beta}} \) and \( \Omega_{\bar{\beta}} \). Then the well known Hausman test statistic is

\[ N \left( \tilde{\beta} - \bar{\beta} \right)' \left( \tilde{\Omega}_{\tilde{\beta}} + \bar{\Omega}_{\bar{\beta}} \right)^{-1} \left( \tilde{\beta} - \bar{\beta} \right), \]

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

3 A Potential Problem with Conventional Procedure

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

\[
\hat{\beta} = \frac{\sum_{i=1}^{N} (x_{i2} - x_{i1}) (y_{i2} - y_{i1})}{\sum_{i=1}^{N} (x_{i2} - x_{i1})^2}
\]

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

\[
\hat{\beta} = \frac{1}{m} \sum_{i=1}^{m} (y_{i2} - y_{i1}).
\]

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.\(^2\)

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

4 A Bootstrap-Like Solution

Consider the moment condition

\[
E \left[ \sum_{i=1}^{T} \bar{x}_{it} (\bar{y}_{it} - \bar{x}_{it}' \hat{\beta}_B) \right] = 0,
\]

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

Define

\[
\hat{\Sigma}_\hat{x} = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \bar{x}_{it} \bar{x}_{it}', \quad \hat{\Sigma}_\bar{y} = \frac{1}{N} \sum_{i=1}^{N} (\bar{x}_i - \bar{\bar{x}}) (\bar{x}_i - \bar{\bar{x}})',
\]

\[
\hat{\sigma}_v^2 = \frac{1}{N(T-1)} \sum_{i=1}^{N} \hat{e}_i \hat{e}_i', \quad \text{and} \quad \hat{\Omega}_\beta = \left( \frac{1}{N} \sum_{i=1}^{N} (\bar{x}_i - \bar{\bar{x}}) (\bar{x}_i - \bar{\bar{x}})' \right)^{-1} \left( \frac{1}{N} \sum_{i=1}^{N} \hat{f}_i^2 \right),
\]

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

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

\(^2\)We discuss this problem further in our Online appendix Hahn, Ham and Moon (2010).
A natural statistic for the moment condition (4) is

\[ H = \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_{it} (\bar{y}_{it} - \bar{x}_{it}'\hat{\beta}) \right)' \left[ \hat{\sigma}_x^2 \hat{\Sigma}_2 + \hat{\Sigma}_x \hat{\Omega}_\beta \hat{\Sigma}_x \right]^{-1} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_{it} (\bar{y}_{it} - \bar{x}_{it}'\hat{\beta}) \right). \]

In fact, it is easy to see that when \( \hat{\Omega}_\beta = \hat{\sigma}_x^2 \hat{\Sigma}_x^{-1} \), 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

\[ \hat{e}_i = \bar{y}_i - \bar{x}_i \hat{\beta} - (\bar{y} - \bar{x} \hat{\beta}) \] \hspace{1cm} (5)
\[ \hat{f}_i = y_i - x_i' \hat{\beta} - (\bar{y} - \bar{x} \hat{\beta}) \] \hspace{1cm} (6)

(Note that we are de-meaning the residuals.)

2. From the empirical distribution \( F_N \) of the sample \( \{(\hat{e}_i, \hat{f}_i)\} \), generate a random sample \( \{(\hat{e}_i^*, \hat{f}_i^*)\} \).

3. Let

\[ H^* = \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_{it} \hat{e}_{it}^* - \hat{\Sigma}_x \hat{\Sigma}_x^{-1} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\bar{x}_i - \bar{x}) f_i^* \right) \right)' \left[ \hat{\sigma}_x^2 \hat{\Sigma}_x + \hat{\Sigma}_x \hat{\Omega}_\beta \hat{\Sigma}_x \right]^{-1} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_{it} \hat{e}_{it}^* - \hat{\Sigma}_x \hat{\Sigma}_x^{-1} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\bar{x}_i - \bar{x}) f_i^* \right) \right). \]

4. Repeat 2 and 3 many times and and tabulate the distribution of \( H^* \).

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 \( \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{x}_{it}' \) may have different rates of convergence. In order to express this idea, we will let \( \delta_k \) denote the rates of convergence of the \((k, k)\) element of \( \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{x}_{it}' \), i.e., we will assume that

\[ \frac{1}{N^\delta_k} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_{k,it}^2 \]

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

\[ \hat{\Lambda} = \Delta_n^{-1} \hat{\Sigma}_x \Delta_n^{-1} \]

\[ = \text{diag} \left( N^{-\delta_1/2}, ..., N^{-\delta_{\text{dim}(\beta)/2}) \right) \left( \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{x}_{it}' \right) \text{diag} \left( N^{-\delta_1/2}, ..., N^{-\delta_{\text{dim}(\beta)/2}) \right). \]
Below we assume that $\hat{\Lambda}$ has a well defined limit $\Lambda$. Let $\Delta = \lim_n \Delta_n$.

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

**Condition 1** (a) $\varepsilon_{it}$ are iid over $i$ and $t$; (b) $E[\varepsilon_{it}] = 0$ and $E[|\varepsilon_{it}|^8] < \infty$; (c) $u_i$ is independent of $(\varepsilon_{i1}, \ldots, \varepsilon_{iT})$; (d) $E[u_i] = 0$ and $E[u_i^8] < \infty$; (e) $(x_{i1}, \ldots, x_{iT})$ 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; and (g) $\limsup \frac{1}{N} \sum_i \|x_i\|^8 < \infty$; (h) $0 \leq \delta_k \leq 1$ for all $k = 1, \ldots, \dim(\beta)$, and $\lim \Lambda = \Lambda$, where $\Lambda$ is a positive definite matrix; (i) $\sigma^2 \Lambda + \Lambda \Delta \Omega_\beta' \Delta \Lambda$ is invertible.

Let $L^0_N$ 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 $\rho\left(L^0_N, L^*_N\right)$ converges to zero in probability, where $\rho(\cdot, \cdot)$ denotes the Prohorov metric.

**Theorem 1** Under Condition 1, $\rho\left(L^0_N, L^*_N\right) = o_p(1)$.

**Proof.** In appendix. ■

Theorem 1 establishes that our bootstrap based procedure $L^*_N$ approximates the $L^0_N$ 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 $\sum_{i=1}^N \sum_{t=1}^T \bar{x}_{k, it}^2$ 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.

**Appendix**

**A Proof of Theorem 1**

Let $\Gamma_\nu = \Gamma_\nu(B)$ be the set of probabilities $\nu$ on a Borel $\sigma$-field of $B$ such that $\int \|z\|^p \nu(dz) < \infty$. For $\nu, \nu^* \in \Gamma_\nu$, 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 Bickel and Freedman (1981, Lemma 8.1), the infimum is attained and $d_p$ is a metric on $\Gamma_\nu$. Write $\bar{x}_i = (\bar{x}_{i1}, \ldots, \bar{x}_{iT})'$ and $\bar{\varepsilon}_i = (\bar{\varepsilon}_{i1}, \ldots, \bar{\varepsilon}_{iT})'$.

We begin with a few lemmas. Proofs of Lemmas 1, 2, and 3 are available in our Online Appendix Hahn, Ham and Moon (2010).
Lemma 1 Under Condition 1, (a) $\left( \beta - \bar{\beta} \right)' \left( \frac{1}{N} \sum_{i=1}^{N} \tilde{x}_i x_i \right) \left( \beta - \bar{\beta} \right) = o_{a.s.}(1)$, and (b) $\left( \beta - \beta_B \right)' \left( \frac{1}{N} \sum_{i=1}^{N} (\bar{x}_i - \bar{x})(\tilde{x}_i - \tilde{x})' \right) \left( \beta - \beta_B \right) = o_{a.s.}(1)$.

Lemma 2 Under Condition 1, $\hat{\sigma}_2^2 = \frac{1}{N(N-1)} \sum_{i=1}^{N} \tilde{e}_i^T \tilde{e}_i = \sigma_2^2 + o_{a.s.}(1)$.

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

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

Proof. Notice that

$$d_2(\psi, \psi^*)^2 = d_2 \left( \left[ \frac{\Delta_n^{-1}}{\sqrt{N}} \sum_{i=1}^{N} \tilde{x}_i \tilde{e}_i \right], \left[ \frac{\Delta_n^{-1}}{\sqrt{N}} \sum_{i=1}^{N} \tilde{x}_i \tilde{e}_i^* \right] \right)^2 \leq \sum_{i=1}^{N} d_2 \left( \left[ \frac{\Delta_n^{-1}}{\sqrt{N}} \tilde{x}_i \tilde{e}_i \right], \left[ \frac{\Delta_n^{-1}}{\sqrt{N}} \tilde{x}_i \tilde{e}_i^* \right] \right)^2$$

by Bickel and Freedman (1981, Lemma 8.7). Denote $\xi_i = (\tilde{x}_i, \tilde{e}_i)'$ and $\hat{\xi}_i^* = (\tilde{x}_i, f_i^*)'$. Use Bickel and Freedman (1981, equations (8.2) and (8.3)) and bound the RHS by

$$O(1) d_2 \left( \xi_i, \hat{\xi}_i^* \right)^2,$$

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

$$d_2 \left( \xi_i^*, \hat{\xi}_i^* \right)^2 = d_2 \left( \xi_i^*, \xi_i \right)^2 + ||\xi||^2 = d_2 \left( \xi_i^*, \xi_i \right)^2 \left( \left[ \frac{\bar{y} - \bar{x}}{\bar{y} - \bar{x}} \right] \right)^2 + ||\xi||^2 = d_2 \left( \xi_i^*, \xi_i \right)^2 \left( \left[ \frac{\bar{y} - \bar{x}}{\bar{y} - \bar{x}} \right] \right)^2 + ||\xi||^2.$$
By definition
\[
d_2 \left( \xi_i^*, \tilde{\xi}_i^* \right) + \left( \frac{y - \bar{x}}{y - \bar{x}} \right) \right)^2 \leq \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{\tilde{\varepsilon}_i - \tilde{y}_i + \tilde{x}_i \beta}{\xi_i - \tilde{y}_i + \tilde{x}_i \beta} \right\|^2
\]
\[
= \left( \beta - \beta \right)^2 \left( \frac{1}{N} \sum_{i=1}^{N} x_i x'_i \right) (\beta - \beta) + (\beta - \beta_B)^2 \left( \frac{1}{N} \sum_{i=1}^{N} x_i x'_i \right) (\beta - \beta_B) = o_{a.s.} (1),
\]
where the last equality is based on Lemma 1. Also,
\[
\left\| \tilde{\varepsilon}_0 - y + x' \beta \right\|^2 \leq \frac{1}{N} \sum_{i=1}^{N} \left\| \tilde{\varepsilon}_i - \tilde{y}_i + \tilde{x}_i \beta \right\|^2 \quad \text{(by the Cauchy-Schwarz inequality)}
\]
\[
= \frac{1}{N} \sum_{i=1}^{N} \left\| \tilde{x}_i (\tilde{\beta} - \beta) \right\|^2 = o_{a.s.} (1),
\]
and likewise, \( \left\| \tilde{\xi}_0 - y + x' \beta \right\| = o_{a.s.} (1) \). Finally, \( \left\| \tilde{\xi} \right\|^2 = \left\| \frac{1}{N} \sum_{i=1}^{N} \tilde{\xi}_i \right\|^2 = o_{a.s.} (1) \) by SLLN. Therefore, we deduce \( d_2 \left( \xi_i^*, \tilde{\xi}_i^* \right) = o_{a.s.} (1) \), and \( d_2 (\psi, \psi^*) = o_{a.s.} (1) \), as required. \( \blacksquare \)

**Proof of Theorem 1**

The ‘denominator’ of \( H \) can be rewritten as \( \tilde{\Omega}_2 \tilde{\Gamma}_2 + \tilde{\Omega}_2 \tilde{\Gamma}_2 = \Delta_n \left( \tilde{\Omega}_2 \Lambda + \tilde{\Lambda} \Delta_n \tilde{\Omega}_2 \Lambda \right) \Delta_n \). By Lemmas 2, 3, and Condition 1(h), we have
\[
\tilde{\Omega}_2 \Lambda + \tilde{\Lambda} \Delta_n \tilde{\Omega}_2 \Lambda = \sigma_2^2 \Lambda + \Lambda \Delta \Lambda + o_{a.s.} (1),
\]
where \( \Delta \equiv \lim \Delta_n \). By definition \( \tilde{\Gamma}_2 = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_i t \tilde{x}_i' \). Let \( H_{null} = \Psi^* \left( \tilde{\Omega}_2 \Lambda + \tilde{\Lambda} \Delta_n \tilde{\Omega}_2 \Lambda \right)^{-1} \Psi \)
where
\[
\Psi \equiv \Delta_n^{-1} \left( \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_i t \tilde{x}_i' \right) - \tilde{\Omega}_2 \tilde{\Gamma}_2^{-1} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{x}_i \tilde{x}_i' \right) \right)
\]
\[
= \Delta_n^{-1} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_i t \tilde{x}_i' \right) - \tilde{\Lambda} \Delta_n \tilde{\Omega}_2^{-1} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{x}_i \tilde{x}_i' \right) - \psi_1 - \tilde{\Lambda} \Delta_n \tilde{\Omega}_2^{-1} \psi_2.
\]
Note that the distribution of \( H_{null} \) is equal to the null distribution \( L_N^0 \) of \( H \). Also, write \( H^* = (\Psi^*)^* \left( \tilde{\Omega}_2 \Lambda + \tilde{\Lambda} \Delta_n \tilde{\Omega}_2 \Lambda \right)^{-1} \Psi^* \), where
\[
\Psi^* \equiv \Delta_n^{-1} \left( \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_i t \tilde{x}_i' \right) - \tilde{\Omega}_2 \tilde{\Gamma}_2^{-1} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{x}_i \tilde{x}_i' \right) \right)
\]
\[
= \Delta_n^{-1} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_i t \tilde{x}_i' \right) - \tilde{\Lambda} \Delta_n \tilde{\Omega}_2^{-1} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{x}_i \tilde{x}_i' \right) - \psi_1 - \tilde{\Lambda} \Delta_n \tilde{\Omega}_2^{-1} \psi_2^*.
\]
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 \( P^* (\cdot \mid Z^N) \) 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
\[
P^* \left( H^* \leq t \mid Z^N \right) = P \left( H_{null} \leq t \right) + o_p (1). \tag{9}
\]
Define \( \rho (P, Q) = \inf \{ \varepsilon > 0 : P (A) \leq Q (A^\varepsilon) + \varepsilon \) for all Borel sets \( A \}, \) where \( Q (A^\varepsilon) \) denotes the \( \varepsilon \)-inflated set of \( A \) (see page 309 of Dudley 1989). For (9), it suffices to show\(^1\) that \( \rho \left( L^0_N, L^*_N \right) = o_p (1) \), for which it is enough to show that for every subsequence \( \rho \left( L^0_N (n'), L^*_N (n') \right) \), there exists a further subsequence \( \rho \left( L^0_N (n''), L^*_N (n'') \right) \) such that \( \rho \left( L^0_N (n''), L^*_N (n'') \right) = o_{a.s.} (1) \). (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
\[
d_2 \left( \psi (n''), \psi^* (n'') \right) = o_{a.s.} (1) \tag{10}
\]
along the subsequence \( n'' \).

Also, notice by definition that
\[
\psi (n'') \Rightarrow \psi (n'). \tag{11}
\]
Furthermore, under Conditions 1 it is easy to show that \( \{ \| \psi (n'') \|^2 \} \) is uniformly integrable, which yields together with (11) the second moment convergence. Then, by Bickel and Freedman (1981, Lemma 8.3), we have
\[
d_2 \left( \psi (n''), \psi (n') \right) = o (1). \tag{12}
\]
Since \( d_2 (\psi^* (n''), \psi (n')) \leq d_2 (\psi (n''), \psi^* (n'')) + d_2 (\psi (n''), \psi (n')) \), from (10) and (12), we deduce \( d_2 (\psi^* (n''), \psi (n')) = o_{a.s.} (1) \). Then, by Bickel and Freedman (1981, Lemma 8.3) we have \( \psi^* (n'') \Rightarrow \psi (n') \). Denote \( \Psi (n') = \psi_1 (n') - \Lambda \Delta_{n} \Sigma_{n}^{-1} \psi_2 (n') \). We now use (8) and the continuous mapping theorem to deduce that
\[
H_{null} = \Psi' \left( \delta^2 \Lambda + \Lambda \Delta_{n} \Sigma_{n} \Lambda \right)^{-1} \Psi
\Rightarrow \Psi' \left( \delta^2 \Lambda + \Lambda \Delta \Sigma_{\infty} \Lambda \right)^{-1} \Psi (n') = L (n') \text{, say}
\]
along the subsequence \( n'' \). Similarly, we can show that \( H^* \Rightarrow L (n') \) a.s. along the subsequence \( n'' \). Then, by Theorem 11.3.3. of Dudley (1989), since \( H_{null} \Rightarrow L (n') \) implies \( \rho \left( H_{null}, L (n') \right) = o (1) \) and \( H^* \Rightarrow L (n') \) a.s. implies \( \rho \left( H^*, L (n') \right) = o_{a.s.} (1) \) along the subsequence \( n'' \) and we can claim that
\[
\rho \left( L^0_N (n''), L^*_N (n'') \right) \leq \rho \left( L^0_N (n''), L (n') \right) + \rho \left( L (n'), L^*_N (n'') \right) = o_{a.s.} (1),
\]
as required. \( \blacksquare \)

\(^1\)We can show the result by modifying the proof of Thereom 11.3.3 of Dudley (1989), Part \( (d) \Rightarrow (a) \).
References


