Test of Random vs Fixed Effects with Small Within Variation

Jinyong Hahn — UCLA
John Ham — Maryland
Hyungsik Rorger Moon — USC

September 3, 2009

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. The modified procedure is robust in the sense that it remains valid whether the within variation is small or large.

1 Introduction

With the advent of many panel data sets[1] 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, but in many instances one sees researchers taking a stand on random versus fixed effects without doing the Hausman test. The problem is that if the within variation is small, the fixed effects estimates may not be asymptotically normal, which may invalidate the basic premise of the Hausman test.

One example of this issue occurs in the recent literature on the effect of welfare reform on the health insurance coverage of single mothers. Welfare reform lead to many single mothers leaving the welfare system and thus giving up their family's Medicaid Health insurance that automatically comes with receiving welfare. The approach taken in most previous nonexperimental studies is to estimate a linear probability model with fixed effects on monthly data and include a dummy variable for TANF

[1] US panel data sets used by researchers include the Panel Study of Income Dynamics, the Survey of Income and Program Participation and the various National Longitudinal Surveys. The main Canadian panel data set is the Survey of Labour and Income Dynamics, while the main British panel data set is the British Household Panel Survey.

adoption (welfare reform) by the state in which the mother leaves. Researchers often have a series of repeated panels (as in the SIPP), only some of which cover the period in which TANF was introduced. As a result the TANF effect is identified from only the relatively small number of individuals who are observed both before and after the introduction of TANF and researchers often assume this is not sufficient to identify the parameter of interest, or to give a Hausman test of random versus fixed effects much power. As a result, they often modify their approach by introducing the possibility of an untreated comparison group. For example, Bitler, Gelbach, and Hoynes (2005) treat single women as the treatment group and married women as the comparison group.[2] Alternatively, Cawley, Schroeder, and Simon (2005) chose single mothers as the group affected by welfare reform (the treatment group) and used married mothers or single women without children as comparison groups. However, for this to be a valid strategy, one must assume that the parameters across the treatment and comparison groups on time varying explanatory variables, including coefficients on any trend terms, are equal across the groups. Ham, Li and Shore-Sheppard (2009) argue that the assumption of common coefficients across treatment and comparison groups made in previous studies is not credible, and instead choose to include a richer set of explanatory variables and use random effects. They then test the assumption of common coefficients across the different groups under random effects estimation, and find that neither married mothers nor single women are a valid comparison group for the treatment group of single women with children. Thus all of these researchers avoided using fixed effects estimation on single mothers, but did not test the assumptions necessary for using an alternative treatment group (as in the case of Bitler, Gelbach and Hoynes or Cawley, Schroeder, and Simon) or using random effects (in the case of Ham, Li and Shore-Sheppard).

The purpose of this paper is largely theoretical. We show that researchers' 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. We then 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 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 iid over i and t and the individual effect

[2] Their data set has the drawback that they cannot observe which single women have children in their data, but, on the other hand, it also has several advantages over more traditional data sets.

parameter is specified to be

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

where u_i is independent of $(x_{i1}, \ldots, x_{iT})$ and $(\varepsilon_{i1}, \ldots, \varepsilon_{iT})$.[3] The 'between' and 'within' models are

$$\begin{aligned} \overline{y}_i &= \alpha_i + \overline{x}_i'\beta + \overline{\varepsilon}_i \text{ and} \\ \widetilde{y}_{it} &= \widetilde{x}_{it}'\beta + \widetilde{\varepsilon}_{it}, \end{aligned}$$

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

$$H_0 : \gamma = 0.$$

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

$$\overline{\beta} = \left(\sum_{i=1}^{N} \overline{x}_i\overline{x}_i'\right)^{-1} \left(\sum_{i=1}^{N} \overline{x}_i\overline{y}_i\right).$$

As noted by Hausman (1978), the comparison of random and fixed effects estimators under conventional asymptotics is equivalent to the comparison of between and within estimators. Letting

$$\beta_B = \operatorname{plim} \overline{\beta} = \beta + \gamma,$$

it is typically shown that

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

for some $\Omega_{\widetilde{\beta}}$ and $\Omega_{\overline{\beta}}$. Therefore, a natural statistic is $N\left(\widetilde{\beta} - \overline{\beta}\right)'\left[\widehat{\Omega}_{\widetilde{\beta}} + \widehat{\Omega}_{\overline{\beta}}\right]^{-1}\left(\widetilde{\beta} - \overline{\beta}\right)$, where $\widehat{\Omega}_{\widetilde{\beta}}$ and $\widehat{\Omega}_{\overline{\beta}}$ are some consistent estimators of population counterparts. Based on asymptotic normality of $\left(\sqrt{N}\left(\widetilde{\beta} - \beta\right), \sqrt{N}\left(\overline{\beta} - \beta_B\right)\right)$, the asymptotic distribution of the test statistic under the null is understood to be $\chi^2_{\dim(\beta)}$ - see Hausman and Taylor (1981).

3 Potential Problem with Conventional Procedure

Implicit in the conventional Hausman test is the assumption that both $\sqrt{N}\left(\widetilde{\beta} - \beta\right)$ and $\sqrt{N}\left(\overline{\beta} - \beta_B\right)$ are asymptotically normal, which in turn requires that both $\widetilde{x}$ and $\overline{x}$ have sufficient variation. Although the between variation (i.e., $\sum_{i=1}^{N} \overline{x}_i\overline{x}_i'$) is typically large, the within variation (i.e., $\sum_{i=1}^{N}\sum_{t=1}^{T} \widetilde{x}_{it}\widetilde{x}_{it}'$)

[3] See, e.g., Mundlak (1978) for the specification under the alternative.

tends to be small in many applications. As a consequence, the asymptotic normality of $\sqrt{N}\left(\widetilde{\beta}-\beta\right)$ 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

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

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

$$\frac{1}{m}\sum_{i=1}^{m}\left(y_{i2}-y_{i1}\right)$$

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 effects estimates are very noisy. This was certainly the case in the papers studying the effect of welfare reform discussed above. Another instance occurs in Sawangfa (2007), who investigates the effect of job changing on job satisfaction using Korean panel data. She finds that fixed effects estimation produces such large confidence intervals for the parameter of interest that are basically uninformative, while she finds that the random effects estimators which also exploit variation across individuals show a positive and statistically significant effect of job changing on job satisfaction. A similar problem will arise in studies examining the effect of schooling or training on earnings if there is little variation (as is usually thought to be the case) within an individual in these variables, or the effect on consumption of a once-and-for-all policy change such as legalized gambling (Kearney 2005).

4 Bootstrap-Like Approach

It is well-known[5] that the conventional procedure can be understood to be a test of moment restriction

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

[4] We discuss this problem further in on-line appendix. The within estimator $\widetilde{\beta}$ may be asymptotically normal (with a slower rate of convergence) even when $\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{x}_{it}'$ is small if some regularity condition is satisfied. The regularity condition is likely to be violated when the regressor is binary, as is the case in many empirical applications.

[5] See Hausman and Taylor (1981).

where $\beta_B = \text{plim}\,\overline{\beta}$ is a solution to the moment equation $E\left[\overline{x}_i\left(\overline{y}_i - \overline{x}_i'b\right)\right] = 0$. A natural statistic is then

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

where

$$\widehat{\Sigma}_{\widetilde{x}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{x}_{it}', \quad \widehat{\Sigma}_{\overline{x}} = \frac{1}{N}\sum_{i=1}^{N}\overline{x}_i\overline{x}_i',$$

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

where $\widehat{\widetilde{e}}_i$ and $\widehat{f}_i$ are defined below in equations (1) and (2). The matrices $\widehat{\Sigma}_{\widetilde{x}}$, and $\widehat{\Omega}_{\overline{\beta}}$ are estimates of $\Sigma_{\widetilde{x}}$, and the asymptotic variance

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

of $\sqrt{N}\left(\overline{\beta} - \beta_B\right)$.

We will obtain the critical value by simulation using the following algorithm:

1. Compute

$$\widehat{\widetilde{e}}_i = \widetilde{y}_i - \widetilde{x}_i\widetilde{\beta} - \left(\overline{\widetilde{y}} - \overline{\widetilde{x}}\widetilde{\beta}\right) \tag{1}$$

$$\widehat{f}_i = \overline{y}_i - \overline{x}_i'\overline{\beta} - \left(\overline{\overline{y}} - \overline{\overline{x}}'\overline{\beta}\right) \tag{2}$$

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

2. From the empirical distribution F_N of the sample $\left\{\left(\widehat{\widetilde{e}}_i, \widehat{f}_i\right)\right\}$, generate a random sample $\{(\widetilde{e}_i^{*}, f_i^{*})\}$.

3. Let

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

and tabulate its distribution.

Remark 1 *Our procedure is not exactly a bootstrap procedure because we use the same "denominator"* $\left(\widehat{\sigma}_{\varepsilon}^{2}\widehat{\Sigma}_{\widetilde{x}} + \widehat{\Sigma}_{\widetilde{x}}\widehat{\Omega}_{\overline{\beta}}\widehat{\Sigma}_{\widetilde{x}}\right)$ *when we calculate the* H^{*}*. This serves a practical purpose. Consider resampling* $w_i = (x_{i1}, \ldots, x_{iT}, y_{i1}, \ldots, y_{iT})$*, and recalculating* $\widehat{\Sigma}_{\widetilde{x}}$*. In the extreme example discussed in the previous* *section, it happens with positive probability that $\widehat{\Sigma}_{\widetilde{x}} = 0$, which renders the H^* undefined. Besides, resampling of w_i is predicated on the idea that w_i is iid, which is not always consistent with the asymptotic approximation where the within variation is allowed to be small. (See our asymptotic theory discussion in the next section.)*

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 rates 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

$$\frac{1}{N^{\delta_k}}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{k,it}^2 = O\left(1\right). \tag{3}$$

Writing

$$\Delta_n = \operatorname{diag}\left(N^{-(1-\delta_1)/2}, ..., N^{-\left(1-\delta_{\dim(\beta)}\right)/2}\right),$$

we can see 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

$$\hat{\Lambda} = \Delta_n^{-1}\widehat{\Sigma}_{\widetilde{x}}\Delta_n^{-1}.$$

Condition 1 below elaborates on (3), and imposes other regularity conditions:

Condition 1 *(a) ε_{it} are iid over i and t; (b) $E\left[\varepsilon_{it}\right] = 0$ and $E\left[\left|\varepsilon_{it}\right|^3\right] < \infty$; (c) u_i is independent of $(\varepsilon_{i1},\ldots,\varepsilon_{iT})$; (d) $E\left[u_i\right] = 0$ and $E\left[\left|u_i\right|^3\right] < \infty$; (e) $(x_{i1},\ldots,x_{iT})$ is a nonstochastic triangular array; (f) $\frac{1}{N}\sum_i \overline{x}_i\overline{x}_i'$ converges to a positive definite limit; and (g) $\limsup \frac{1}{N}\sum_i \left\|\overline{x}_i\right\|^3 < \infty$; (h) $0 \le \delta_k \le 1$ for all $k = 1,\ldots,\dim\left(\beta\right)$, and $\lim\hat{\Lambda} = \Lambda$, where Λ is a positive definite matrix.*

Let L_0 denote the distribution of H under the null. Also, let L_* denote the distribution of H^* (given the sample). Then, we show that the distribution of H^* is close to that of the null distribution of H. This is done by showing that $\rho\left(L_0, L_*\right)$ converges to zero, where $\rho\left(\cdot,\cdot\right)$ denotes the Prohorov metric[6]:

Theorem 1 *Under Condition 1, $\rho\left(L_0, L_*\right) \to 0$ almost surely.*

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

[6] Prohorov metric metrizes weak convergence on any separable metric space such as Euclidean space. See, e.g., Dudley (1989, Theorem 11.3.3).

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 discussed in Section 3.

6 Summary

In this paper, we noted that comparison of within and between estimators using Hausman test can be potentially problematic, especially when the within variation is not large enough. This problem occurs because the conventional Hausman test is predicated on asymptotic normality of both estimators, but the distribution of the within estimator may not be approximately normal if the within variation is not sufficiently large. We propose a modification of the Hausman test, and discuss its property under various asymptotic approximations. It is shown that the modified procedure is valid whether the within variation is large or small.

Appendix

A Auxiliary Results

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 $Z^* \sim \nu^*$. The following results from Bickel and Freedman (1981, BF81 hereafter) are useful.

Lemma 1 (BF81, Lemma 8.1) *(a) The infimum is attained; (b) d_p is a metric on Γ_p.*

Lemma 2 (BF81, Lemma 8.3) *Let $\nu_n, \nu \in \Gamma_p$. Then $d_p(\nu_n, \nu) = o(1)$ is equivalent to each of the following: (a) $\nu_n \to \nu$ weakly and $\int \|z\|^p \nu_n(dz) \to \int \|z\|^p \nu(dz)$; (b) $\nu_n \to \nu$ weakly and $\|z\|^p$ is uniformly ν_n-integrable; and (c) $\int \phi \mathrm{d}\nu_n \to \int \phi \mathrm{d}\nu$ for every continuous ϕ such that $\phi(z) = 0(\|z\|^p)$ at infinity.*

Lemma 3 (BF81, Lemma 8.4) *Let Z_i be independent B-valued random variables, with common distribution $\mu \in \Gamma_p$. Let μ_n be the empirical distribution of $Z_1, \ldots, Z_n$. Then $d_p(\mu_n, \mu) \to 0$ almost surely.*

Definition 1 *For B-valued random variables Z and Z^*, we abuse notation slightly and write $d_p(Z, Z^*)$ for the d_p-distance between the probability distributions of Z and Z^*.*

Lemma 4 (BF81, equations (8.2) and (8.3)) *$d_p(aZ, aZ^*) = |a| \cdot d_p(Z, Z^*)$ and $d_p(LZ, LZ^*) = \|L\| \cdot d_p(Z, Z^*)$ for any linear operator L on B.*

Lemma 5 (BF81, Lemma 8.7) *Suppose B is a Hilbert space with inner product $\langle \cdot, \cdot \rangle$, and $p = 2$. Suppose the U_i are independent, likewise for V_i; assume that the probability distributions are in Γ_2, and that $E[U_j] = E[V_j]$. Then*

$$d_2\left(\sum_{i=1}^m U_i, \sum_{i=1}^m V_i\right)^2 \leq \sum_{i=1}^m d_2(U_i, V_i)^2.$$

Lemma 6 (BF81, Lemma 8.8) *For two random vectors U and V with $E\|U\|^2 < \infty$ and $E\|V\|^2 < \infty$,*

$$d_2(U, V)^2 = d_2(U - E(U), V - E(V))^2 + \|E(U) - E(V)\|^2.$$

B Proof of Theorem 1

Before we start the proof of Theorem 1, we introduce a few lemmas: Write $\widetilde{x}_i = (\widetilde{x}_{i1}, ..., \widetilde{x}_{iT})'$ and $\widetilde{\varepsilon}_i = (\widetilde{\varepsilon}_{i1}, ..., \widetilde{\varepsilon}_{iT})'$.

Lemma 7 *Under Condition 1,*

$$\left(\widetilde{\beta} - \beta\right)' \left(\frac{1}{N}\sum_{i=1}^{N} \widetilde{x}_i' \widetilde{x}_i\right) \left(\widetilde{\beta} - \beta\right) = o_{a.s}(1),$$

and

$$\left(\overline{\beta} - \beta_B\right)' \left(\frac{1}{N}\sum_{i=1}^{N} \overline{x}_i \overline{x}_i'\right) \left(\overline{\beta} - \beta_B\right) = o_{a.s}(1).$$

Proof. Because $\widetilde{\varepsilon}_i$ are iid with finite variance across i, it follows that

$$\left(\sum_{i=1}^{N} \widetilde{x}_i' \widetilde{x}_i\right)^{1/2} \left(\widetilde{\beta} - \beta\right) = \sum_{i=1}^{N} \left(\sum_{i=1}^{N} \widetilde{x}_i' \widetilde{x}_i\right)^{-1/2} \widetilde{x}_i' \widetilde{\varepsilon}_i$$

is an L_2 – martingale with respect to the natural filteration. Also notice that

$$\begin{aligned}
\sup_N E\left\| \sum_{i=1}^{N} \left(\sum_{i=1}^{N} \widetilde{x}_i' \widetilde{x}_i\right)^{-1/2} \widetilde{x}_i' \widetilde{\varepsilon}_i \right\|^2 &= \sup_N E\left[\operatorname{trace}\left(\left(\widetilde{\beta} - \beta\right)' \left(\sum_{i=1}^{N} \widetilde{x}_i' \widetilde{x}_i\right) \left(\widetilde{\beta} - \beta\right)\right)\right] \\
&= \sup_N \operatorname{trace}\left\{\left(\sum_{i=1}^{N} \widetilde{x}_i' \widetilde{x}_i\right) \cdot \sigma_\varepsilon^2 \left(\sum_{i=1}^{N} \widetilde{x}_i' \widetilde{x}_i\right)^{-1}\right\} \\
&= \dim(\beta) = K < \infty..
\end{aligned}$$

Then, by the martingale convergence theorem (e.g., Theorem 10.5.4 in Dudley(1989)), the martingale $\left(\sum_{i=1}^{N} \widetilde{x}_i' \widetilde{x}_i\right)^{1/2} \left(\widetilde{\beta} - \beta\right)$ converges almost surely to a finite limit random variable, from which we deduce the required result for the lemma. The second claim follows easily from $\overline{\beta} = \beta_B + o_{a.s.}(1)$.[7] ■

Lemma 8 *Under Condition 1,* $\widehat{\sigma}_\varepsilon^2 = \frac{1}{N(T-1)} \sum_{i=1}^{N} \widehat{\widetilde{e}}_i' \widehat{\widetilde{e}}_i = \sigma_\varepsilon^2 + o_{a.s}(1)$.

Proof. Note first that

$$\frac{1}{N}\sum_{i=1}^{N} \left(\widetilde{\varepsilon}_i - \overline{\widetilde{\varepsilon}}\right)' \left(\widetilde{\varepsilon}_i - \overline{\widetilde{\varepsilon}}\right) = \sigma_\varepsilon^2 + o_{a.s}(1)$$

[7] See Lai, Robbins, and Wei (1979).

by the strong law of large numbers. Also note that

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\left(\widetilde{x}_{it}-\overline{\widetilde{x}}_{t}\right)'\left(\widetilde{\beta}-\beta\right)\right)^{2} &= \left(\widetilde{\beta}-\beta\right)'\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\widetilde{x}_{it}-\overline{\widetilde{x}}_{t}\right)\left(\widetilde{x}_{it}-\overline{\widetilde{x}}_{t}\right)'\right)\left(\widetilde{\beta}-\beta\right) \\
&\leq \left(\widetilde{\beta}-\beta\right)'\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{x}_{it}'\right)\left(\widetilde{\beta}-\beta\right) \\
&= \left(\widetilde{\beta}-\beta\right)'\left(\frac{1}{N}\sum_{i=1}^{N}\widetilde{x}_{i}'\widetilde{x}_{i}\right)\left(\widetilde{\beta}-\beta\right) \\
&= o_{a.s.}(1),
\end{aligned}$$

where the last equality is based on Lemma 7. These two facts, combined with the equality $\widehat{\widetilde{e}}_i = \widetilde{\varepsilon}_i-\overline{\widetilde{\varepsilon}}-\left(\widetilde{x}_i-\overline{\widetilde{x}}\right)\left(\widetilde{\beta}-\beta\right)$, imply the desired conclusion. ■

Lemma 9 *Under Condition 1,* $\widehat{\Omega}_{\overline{\beta}} = \Omega_{\overline{\beta}} + o_{a.s.}(1)$.

Proof. This follows from $\overline{\beta} = \beta_B + o_{a.s.}(1)$, and hence the details are omitted. ■

Lemma 10 *Let*

$$\begin{aligned}
&\psi_1 \equiv \Delta_n^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{\varepsilon}_{it}\right), \quad \psi_1^* \equiv \Delta_n^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{e}_{it}^*\right), \\
&\psi_2 \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{x}_i\zeta_i, \quad \text{and} \quad \psi_2^* \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{x}_i f_i^*,
\end{aligned}$$

where $\zeta_i \equiv u_i + \overline{\varepsilon}_i$*. 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

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

by Lemma 5. Because $\widetilde{\varepsilon}_i$ and $\widetilde{e}_i^*$ are iid, we can use Lemma 4 and bound the RHS by

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

where

$$\xi_i = \begin{bmatrix} \widetilde{\varepsilon}_i \\ \zeta_i \end{bmatrix} \quad \text{and} \quad \widehat{\xi}_i^* = \begin{bmatrix} \widetilde{e}_i^* \\ f_i^* \end{bmatrix}.$$

For the required result, it remains to prove that

$$d_2\left(\xi_i, \widehat{\xi}_i^*\right) = o_{a.s.}(1).$$

Let $\{\xi_i^*\}_{i=1,\dots,N}$ denote the iid samples from the empirical distribution of $\{\xi_i\}_{i=1,\dots,N}$. Then, by the triangle inequality,

$$d_2\left(\xi_i, \widehat{\xi}_i^*\right) \le d_2\left(\xi_i, \xi_i^*\right) + d_2\left(\xi_i^*,\ \widehat{\xi}_i^*\right).$$

By Lemma 3,

$$d_2\left(\xi_i, \xi_i^*\right) = o_{a.s.}(1).$$

Next, we apply Lemma 6 twice to obtain

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

By definition

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

where the last equality is based on Lemma 7. Also,

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

and likewise, $\left\|\overline{\zeta} - \overline{\overline{y}} + \overline{\overline{x}}'\overline{\beta}\right\| = o_{a.s.}(1)$. Finally,

$$\left\|\overline{\xi}\right\|^2 = \left\|\frac{1}{N}\sum_{i=1}^{N}\xi_i\right\|^2 = o_{a.s.}(1) \text{ by SLLN.}$$

Therefore, we deduce $d_2\left(\xi_i^*,\ \widehat{\xi}_i^*\right) = o_{a.s.}(1)$, and

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

as required. ■

We now return to the proof of Theorem 1. The 'denominator' of H can be rewritten as

$$\widehat{\sigma}_\varepsilon^2\widehat{\Sigma}_{\widetilde{x}} + \widehat{\Sigma}_{\widetilde{x}}\widehat{\Omega}_{\overline{\beta}}\widehat{\Sigma}_{\widetilde{x}} = \Delta_n\left(\widehat{\sigma}_\varepsilon^2\widehat{\Lambda} + \widehat{\Lambda}\Delta_n\widehat{\Omega}_{\overline{\beta}}\Delta_n\widehat{\Lambda}\right)\Delta_n.$$

By Lemmas 8, 9, and Condition 1 (h), we have

$$\widehat{\sigma}_\varepsilon^2\widehat{\Lambda} + \widehat{\Lambda}\Delta_n\widehat{\Omega}_{\overline{\beta}}\Delta_n\widehat{\Lambda} = \sigma_\varepsilon^2\Lambda + \Lambda\Delta\Omega_{\overline{\beta}}\Delta\Lambda + o_{a.s.}(1), \tag{4}$$

where $\Delta \equiv \lim \Delta_n$.

By definition

$$\widehat{\Sigma}_{\widetilde{x}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{x}_{it}' = \Delta_n\widehat{\Lambda}\Delta_n.$$

Define

$$H_N = \Psi'\left(\widehat{\sigma}_\varepsilon^2\widehat{\Lambda} + \widehat{\Lambda}\Delta_n\widehat{\Omega}_{\overline{\beta}}\Delta_n\widehat{\Lambda}\right)^{-1}\Psi,$$

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_N is equal to the null distribution L_0 of H.

Also, write

$$H^* = (\Psi^*)'\left(\widehat{\sigma}_\varepsilon^2\widehat{\Lambda} + \widehat{\Lambda}\Delta_n\widehat{\Omega}_{\overline{\beta}}\Delta_n\widehat{\Lambda}\right)^{-1}\Psi^*,$$

where

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

Note that the distribution of H^* is equal to L_*.

The conclusion of the theorem follows by combining (4) and Lemma 10, using the Continuous Mapping Theorem and applying Lemma 2.

References

[1] Bickel, P.J. and D.A. Freedman (1981): "Some Asymptotic Theory for the Bootstrap", *Annals of Statistics* 9: 1196–1217.

[2] Bitler, M. P., J. B. Gelbach, and H. W. Hoynes (2005): "Welfare Reform and Health.", *Journal of Human Resources* 40: 309-334.

[3] Cawley, J. H., M. Schroeder, and K. I. Simon (2005): "Welfare Reform and the Health Insurance Coverage of Women and Children." *Forum for Health Economics & Policy, Forum: Frontiers in Health Policy Research* 8. (http://www.bepress.com/fhep/8/5.).

[4] Dudley, R.M. (1989): *Real Analysis and Probability.* Pacific Grove, Wadsworth & Brooks/Cole.

[5] Ham, J.C., X. Li and L.S. Shore-Sheppard (2009): "A Re-Examination of the Impact of Welfare Reform on Health Insurance Among Less-Skilled Women." Forthcoming in *Welfare Reform and its Long-Term Consequences for America's Poor*, James P. Ziliak (ed.), Cambridge, UK: Cambridge University Press.

[6] Hausman, J.A. (1978): "Specification Tests in Econometrics," *Econometrica* 46: 1273-1291.

[7] Hausman, J.A., and W.E. Taylor (1981): "Panel Data and Unobservable Individual Effects", *Econometrica* 49: 1377–1398.

[8] Kearney, M.S. (2005): "State Lotteries and Consumer Behavior", *Journal of Public Economics* 89: 2269-2299.

[9] Lai, T.L., H. Robbins, and C.Z. Wei (1979): "Strong Consistency of Least Squares Estimates in Multiple Regression II", *Journal of Multivariate Analysis* 9: 343–361.

[10] Sawangfa, O. (2007): "Job-switching and Job Satisfaction in South Korea", mimeo, USC Department of Economics.

[11] Mundlak, Y. (1978): "On the Pooling of Time Series and Cross Section Data", *Econometrica* 46: 69–85.

Test of Random vs Fixed Effects with Small Within Variation: Online Appendix

Normality of the Within Estimator

Jinyong Hahn, John Ham, & Hyungsik Roger Moon

For simplicity of notation, we will assume that x is a scalar and $T = 2$. The within estimator has the characterization

$$\widetilde{\beta} - \beta = \left(\sum_{i=1}^{N} \widetilde{x}_i^2 \right)^{-1} \sum_{i=1}^{N} \widetilde{x}_i \widetilde{\varepsilon}_i.$$

We will assume that $\sum_{i=1}^{N} \widetilde{x}_i^2 = o(N)$. For simplicity, assume that $E\left[\widetilde{\varepsilon}_i^2\right] = 1$. Let

$$Y_{N,i} = \left(\sum_{i=1}^{N} \widetilde{x}_i^2 \right)^{-1/2} \widetilde{x}_i \widetilde{\varepsilon}_i.$$

Then,

$$E\left[Y_{N,i}\right] = 0 \text{ and}$$
$$E\left[\sum_{i=1}^{N} Y_{N,i}^2\right] = 1.$$

Then, according to the Lindeberg-Feller (LF) CLT,

$$\left(\sum_{i=1}^{N} \widetilde{x}_i^2 \right)^{1/2} \left(\widetilde{\beta} - \beta \right) \Rightarrow \mathcal{N}(0,1),$$

if

$$E\left[\sum_{i=1}^{N} Y_{N,i}^2 1\left\{|Y_{N,i}| > \eta\right\}\right] = \sum_{i=1}^{N} E\left[\frac{\widetilde{x}_i^2 \widetilde{\varepsilon}_i^2}{\sum_{i=1}^{N} \widetilde{x}_i^2} 1\left\{|\widetilde{x}_i \widetilde{\varepsilon}_i| > \left(\sum_{i=1}^{N} \widetilde{x}_i^2 \right)^{1/2} \eta\right\}\right] \to 0 \tag{5}$$

for any $\eta > 0$. In our case, a well known sufficient condition for (5) is the Lyapounov condition

$$E \sum_{i=1}^{N} |Y_{N,i}|^3 \to 0. \tag{6}$$

In the standard case where

$$\frac{1}{N} \sum_{i=1}^{N} \widetilde{x}_i^2 \to_p C > 0,$$

and

$$\frac{1}{N} \sum_{i=1}^{N} |\widetilde{x}_i|^3 = O_p(1) \text{ and } E\left[|\widetilde{\varepsilon}_i|^3\right] < \infty,$$

conditional on $\widetilde{x}_i$, we have

$$\begin{aligned}
&\sum_{i=1}^{N} E\left[\frac{\widetilde{x}_i^2\widetilde{\varepsilon}_i^2}{\sum_{i=1}^{N}\widetilde{x}_i^2} 1\left\{|\widetilde{x}_i\widetilde{\varepsilon}_i| > \left(\sum_{i=1}^{N}\widetilde{x}_i^2\right)^{1/2}\eta\right\}\right] \\
&\leq \frac{1}{\eta}\sum_{i=1}^{N} E\left[\frac{|\widetilde{x}_i|^3\,|\widetilde{\varepsilon}_i|^3}{\left(\sum_{i=1}^{N}\widetilde{x}_i^2\right)^{3/2}}\right] \\
&= \frac{E\,|\widetilde{\varepsilon}_i|^3}{\eta}\frac{\sum_{i=1}^{N}|\widetilde{x}_i|^3}{\left(\sum_{i=1}^{N}\widetilde{x}_i^2\right)^{3/2}} \\
&= \frac{E\,|\widetilde{\varepsilon}_i|^3}{\eta}\frac{\frac{1}{N}\sum_{i=1}^{N}|\widetilde{x}_i|^3}{\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\widetilde{x}_i^2\right)^{3/2}} = O_p\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}$$

Thus, conditional on $\widetilde{x}_i$, the LF condition is satisfied and

$$\left(\sum_{i=1}^{N}\widetilde{x}_i^2\right)^{1/2}\left(\widetilde{\beta}-\beta\right) \Rightarrow \mathcal{N}(0,1).$$

Since the limit does not depends on $\widetilde{x}_i$, the CLT follows unconditionally.

In our case, we have $\sum_{i=1}^{N}\widetilde{x}_i^2 = o(N)$. Even then, the CLT will follow as long as

$$\frac{\sum_{i=1}^{N}|\widetilde{x}_i|^3}{\left(\sum_{i=1}^{N}\widetilde{x}_i^2\right)^{3/2}} \to 0.$$

Then,

$$\begin{aligned}
&\sum_{i=1}^{N} E\left[\frac{\widetilde{x}_i^2\widetilde{\varepsilon}_i^2}{\sum_{i=1}^{N}\widetilde{x}_i^2} 1\left\{|\widetilde{x}_i\widetilde{\varepsilon}_i| > \left(\sum_{i=1}^{N}\widetilde{x}_i^2\right)^{1/2}\eta\right\}\right] \\
&\leq \frac{1}{\eta}\sum_{i=1}^{N} E\left[\frac{|\widetilde{x}_i|^3\,|\widetilde{\varepsilon}_i|^3}{\left(\sum_{i=1}^{N}\widetilde{x}_i^2\right)^{3/2}}\right] \\
&= \frac{E\,|\widetilde{\varepsilon}_i|^3}{\eta}\frac{\sum_{i=1}^{N}|\widetilde{x}_i|^3}{\left(\sum_{i=1}^{N}\widetilde{x}_i^2\right)^{3/2}} \to 0.
\end{aligned}$$

and

$$\left(\sum_{i=1}^{N}\widetilde{x}_i^2\right)^{1/2}\left(\widetilde{\beta}-\beta\right) \Rightarrow \mathcal{N}(0,1).$$

Such condition is expected to be violated when x_{it} is a dummy variable and there are very few observations where the binary regressors change over time. Suppose that the number of such observations is fixed at r. Then, since $\widetilde{x}_i \in \{-1,0,1\}$,

$$\sum_{i=1}^{N}\widetilde{x}_i^2 = r, \quad \sum_{i=1}^{N}|\widetilde{x}_i|^3 = r.$$

Then,

$$\frac{\sum_{i=1}^{N} |\widetilde{x}_i|^3}{\left(\sum_{i=1}^{N} \widetilde{x}_i^2\right)^{3/2}} = \frac{r}{r} = 1.$$