Pseudo-Gaussian and Rank-Based Optimal Tests for Random Individual Effects in Large n Small T Panels

Nezar Bennala[a], Marc Hallin[b,c,d,*]
and
Davy Paindaveine[b,e,*]

[a]Facultés universitaires Saint-Louis
[b]Institut de Recherche en Statistique, E.C.A.R.E.S., Université Libre de Bruxelles
[c]Académie Royale de Belgique
[d]ORFE, Princeton University, and CentER, Tilburg University
[e]Département de Mathématique, Université Libre de Bruxelles.

Abstract

We consider the problem of detecting unobserved heterogeneity, that is, the problem of testing the absence of random individual effects in a $n \times T$ panel. We establish a local asymptotic normality property—with respect to intercept, regression coefficient, the scale parameter σ of the error, and the scale parameter σ_u of individual effects (which is the the parameter of interest)—for given (scaled) density f_1 of the error terms, when n tend to infinity and T is fixed. This result allows, via the Hájek representation theorem, for developing asymptotically optimal rank-based tests for the null hypothesis $\sigma_u = 0$ (absence of individual effects). These tests are locally asymptotically optimal at correctly specified innovation densities f_1, but remain valid irrespective of the actual underlying density. The limiting distribution of our test statistics is obtained both under the null and under sequences of contiguous alternatives. A local asymptotic linearity property is established in order to control for the effect of substituting estimators for nuisance parameters. The asymptotic relative efficiencies of the proposed procedures with respect to the corresponding pseudo-Gaussian parametric tests are derived. In particular, the van der Waerden version of our rank-based tests uniformly dominates, from the point of view of Pitman efficiency, the classical Honda test. Small-sample performances are investigated via a Monte-Carlo study, and confirm theoretical findings.

Key words and phrases : Random effects, panel data, rank tests, local asymptotic normality.

JEL Classification : C12, C14, C21, C23.

*The research of Marc Hallin is supported by the Sonderforschungsbereich "Statistical modelling of nonlinear dynamic processes" (SFB 823) of the Deutsche Forschungsgemeinschaft, and by a Discovery Grant of the Australian Research Council. The research of Davy Paindaveine is supported by an A.R.C. contract of the Communauté Française de Belgique. Marc Hallin and Davy Paindaveine are also members of ECORE, the association between CORE and ECARES.

1 Introduction.

1.1 Testing for random effects in panel data.

Panel data consist of a series of T observations made through time over a number n of cross-sectional items or experimental units. By combining cross-sectional and time-series features, panel data methods are able to identify and estimate effects that are not detectable via pure cross-sectional or time-series methods, while controlling for individual heterogeneity and taking into account dynamics; see Wooldridge (2002), Arellano (2003), Hsiao (2003) or Baltagi (2008) for background reading.

Throughout, we consider the model (Baltagi 2008, Chapter 4)

$$Y_{it} = \mu + \boldsymbol{\beta}'\mathbf{x}_{it} + \nu_{it} \qquad i = 1, \ldots, n, \quad t = 1, \ldots, T \tag{1.1}$$

(arguably the most fundamental one in the area of panel data analysis), with i denoting individuals and t denoting time; Y_{it} is the observed response for individual i at time t, $\mathbf{x}_{it}$ is a vector of K nonstochastic exogenous regressors, and $(\mu, \boldsymbol{\beta}')'$ is a $(K+1)\times 1$ unknown regression parameter. The disturbances ν_{it} are assumed to follow the one-way error component model

$$\nu_{it} = u_i + \varepsilon_{it}, \tag{1.2}$$

where u_i denotes unobserved individual-specific random effect and ε_{it} is the error term. We assume that the u_i's are i.i.d. $(0, \sigma_u^2)$, that the ε_{it}'s are i.i.d. $(0, \sigma^2)$, and that u_i and ε_{jt} are mutually independent for all i, j, and t. Note that the individual effects u_i are constant over time, and induce autocorrelation among the data, with $\mathrm{corr}(\nu_{it}, \nu_{is}) = \sigma_u^2/(\sigma_u^2 + \sigma^2)$ for $t \neq s$. Model (1.1) is closer to repeated measurement than to time series models, though, as $\mathrm{corr}(\nu_{it}, \nu_{is})$ remains the same irrespective of the time lag $t - s$: for given i, the ν_{it}'s thus are not mixing nor weakly dependent in any sense. Moreover, autocorrelation is the same for all individuals.

Failing to take that autocorrelation structure into account when estimating the model, e.g. by ordinary least squares (OLS), may lead to seriously biased estimators and invalid testing procedures (Scott and Holt 1982; Moulton 1986). Therefore, it is important to be able to perform a preliminary test of the hypothesis that the variance component σ_u^2 for individual effect is zero, that is, $\mathcal{H}_0 : \sigma_u^2 = 0$, versus the alternative $\mathcal{H}_1 : \sigma_u^2 > 0$.

A popular method for deriving locally optimal tests is the Lagrange Multiplier method based on Gaussian likelihoods. That approach has been considered by Breusch and Pagan (1980), who propose a test of $\mathcal{H}_0$ based on the ("two-sided") Lagrange Multiplier test statistic,

$$T^{(n)}_{\text{Breusch-Pagan}} := \frac{nT}{2(T-1)} \left[\sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{1=l\neq t}^{T} \varepsilon_{it}^{\text{OLS}}\, \varepsilon_{il}^{\text{OLS}} \Big/ \sum_{i=1}^{n} \sum_{t=1}^{T} (\varepsilon_{it}^{\text{OLS}})^2 \right]^2,$$

where $\varepsilon_{it}^{\text{OLS}}$ denotes the OLS residuals. Breusch and Pagan show that this statistic, under $\mathcal{H}_0$, is asymptotically χ_1^2 as both $n \to \infty$ and $T \to \infty$. Honda (1985) and King and Evans (1986) observed that a "one-sided" Lagrange Multiplier test, based on the asymptotically standard normal distribution of

$$T^{(n)}_{\text{Honda}} := \sqrt{\frac{nT}{2(T-1)}} \left[\sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{1=l\neq t}^{T} \varepsilon_{it}^{\text{OLS}}\, \varepsilon_{il}^{\text{OLS}} \Big/ \sum_{i=1}^{n} \sum_{t=1}^{T} (\varepsilon_{it}^{\text{OLS}})^2 \right],$$

is sizeably more powerful in this inherently one-sided problem. Note that the quantity in brackets, which is the same in $T^{(n)}_{\text{Honda}}$ as in $T^{(n)}_{\text{Breusch-Pagan}}$, aims at detecting the autocorrelation induced by the presence of eventual individual effects.

Honda (1985) (for $n \to \infty$ and $T \to \infty$) and, more recently, Orme and Yamagata (2006) (for $n \to \infty$ and fixed T, but assuming, for the ε_{it}'s, finite moments of order $4+\delta$, $\delta > 0$) also established the robustness to nonnormal errors of the asymptotic normality of $T^{(n)}_{\text{Honda}}$. Unfortunately, as shown by Moulton and Randolph (1989), the finite-sample performance of tests based on that normal approximation can be pretty poor, even in fairly large samples. This occurs if either there are many regressors, or the regressors are highly correlated within groups. They suggest a modification of the standardized Lagrange Multiplier statistic, whose asymptotic critical values are generally much closer to the exact critical values than those of the original one. An alternative derivation of the Breusch and Pagan statistic has been obtained by Chesher (1984). Lagrange multiplier tests also have been derived by Hamerle (1990), Orme (1993), Jacqmin-Gadda and Commenges (1995), and Lin (1997) for generalized linear model versions of (1.1).

In practice, small T values, however, are the rule rather than the exception in this context. In a fixed-T setup, Wooldridge (2002) more recently proposed the test statistic

$$T^{(n)}_{\text{Wooldridge}} := \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{1=l\neq t}^{T} \varepsilon^{\text{OLS}}_{it}\,\varepsilon^{\text{OLS}}_{il} \Big/ \left[\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\sum_{1=l\neq t}^{T} \varepsilon^{\text{OLS}}_{it}\,\varepsilon^{\text{OLS}}_{il}\right)^2\right]^{1/2},$$

which is asymptotically standard normal (as $n \to \infty$ under fixed T).

The resulting test, contrary to Honda's, remains asymptotically valid in the presence of temporal heteroskedasticity, that is, when the ε_{it}'s have a variance $\text{Var}(\varepsilon_{it}) = \sigma^2_t$ that may change over time (but remains cross-sectionally constant). If that type of heteroskedasticity is assumed, a better estimator of the variance of the numerator of Wooldridge's statistic can be obtained, though, yielding a modified Wooldridge test statistic of the form

$$T^{(n)}_{\text{Wooldridge}^*} := \sqrt{\frac{n}{2}}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{1=l\neq t}^{T} \varepsilon^{\text{OLS}}_{it}\,\varepsilon^{\text{OLS}}_{il} \Big/ \left[\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{1=l\neq t}^{T} (\varepsilon^{\text{OLS}}_{it})^2(\varepsilon^{\text{OLS}}_{jl})^2\right]^{1/2}.$$

If heteroskedasticity is to be feared, however, a fully general form, where $\text{Var}(\varepsilon_{it}) = \sigma^2_{it}$, can be dealt with by a test proposed by Häggström and Laitila (2002), based on

$$T^{(n)}_{\text{Häggström-Laitila}} := \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{1=l\neq t}^{T} \varepsilon^{\text{OLS}}_{it}\,\varepsilon^{\text{OLS}}_{il} \Big/ \left[2\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{1=l\neq t}^{T} (\varepsilon^{\text{OLS}}_{it})^2(\varepsilon^{\text{OLS}}_{il})^2\right]^{1/2}.$$

This test statistic is shown to be asymptotically standard normal under $\mathcal{H}_0$ ($n \to \infty$, T fixed) provided that some mild assumptions are satisfied, among which finite fourth-order moments for the ε_{it}'s.

While none of the previous tests require specifying the distribution of the random effects u_i, they still are of a parametric nature, and require finite moments, at least of order 2, often of order 4 or stricly larger, for the ε_{it}'s. One way of avoiding such assumptions consists in basing the tests on statistics that are measurable with respect to invariant or distribution-free quantities such as ranks. The main objective of this paper is to construct, for the problem of testing the null hypothesis $\mathcal{H}_0$ of no random effects, rank-based tests which are asymptotically (as $n \to \infty$, under fixed T) distribution-free, hence remain (asymptotically) valid under extremely mild assumptions—in particular, in the absence of any moment assumptions—while retaining optimality (in the Le Cam sense) at correctly specified densities; increasing validity at the expense of efficiency indeed hardly can be considered an improvement.

For instance, the normal-score or van der Waerden rank test proposed in Section 4.3 rejects the null hypothesis of no random individual effects for large values of

$$\underset{\sim}{T}^{*(n)}_{\rm vdW}(\hat{\boldsymbol\beta}) := \frac{a}{2s^{(n)}_{\phi_1}\sqrt{n}} \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{l\neq t=1}^{T} \left\{\Phi^{-1}\Big(\frac{R^{(n)}_{it}(\hat{\boldsymbol\beta})}{N+1}\Big)\Phi^{-1}\Big(\frac{R^{(n)}_{il}(\hat{\boldsymbol\beta})}{N+1}\Big) - c^{(n)}_{\phi_1}\right\},$$

where Φ is the standard normal distribution function, $\hat{\boldsymbol\beta}$ an appropriate estimator of $\boldsymbol\beta$ (OLS are fine, but more robust estimators also can be used), $R^{(n)}_{it}(\boldsymbol\beta)$ stands for the rank of $Z_{it} := Y_{it}-\mu-\boldsymbol\beta'\mathbf{x}_{it}$ among $Z_{11},\ldots,Z_{nT}$ (this rank does not depend on μ, which justifies the notation), $a \approx 0.4549$ (see (2.2)), $c^{(n)}_{\phi_1}$ and $s^{(n)}_{\phi_1}$ are the exact centering and scaling constants defined in (4.5) and (4.2), respectively. This test is (asymptotically, since $\boldsymbol\beta$ is to be replaced with an estimator $\hat{\boldsymbol\beta}$) distribution-free under the hypothesis of no random effects, asymptotically optimal against Gaussian local alternatives, and asymptotically equivalent (as $n\to\infty$, under fixed T) to the Honda test under Gaussian densities.

Although asymptotic relative efficiency (ARE) comparisons between this van der Waerden test and Honda's classical one are somewhat unfair to van der Waerden (since $T^{(n)}_{\rm Honda}$ requires more stringent distributional assumptions than $\underset{\sim}{T}^{*(n)}_{\rm vdW}$), the AREs (see Section 4.4) of $\underset{\sim}{T}^{*(n)}_{\rm vdW}$-based tests with respect to the $T^{(n)}_{\rm Honda}$-based ones are uniformly larger than one. These theoretical findings are confirmed (Section 5) by finite-sample simulations.

A Wilcoxon-type test similarly can be based on the somewhat simpler statistic

$$\underset{\sim}{T}^{*(n)}_{\rm W}(\hat{\boldsymbol\beta}) := \frac{b}{2s^{(n)}_{\ell_1}\sqrt{n}} \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{1=l\neq t}^{T} \left(\frac{R^{(n)}_{it}(\hat{\boldsymbol\beta}_{\#})}{N+1}\frac{R^{(n)}_{il}(\hat{\boldsymbol\beta}_{\#})}{N+1} - c^{(n)}_{\ell_1}\right),$$

with $b = (\ln 3)^2 \approx 1.207$, and exact centering $c^{(n)}_{\ell_1}$ and scale $s^{(n)}_{\ell_1}$; see Section 4.3 for details.

The rank tests we are constructing are alternatives to the Honda's classical test; they are not devised to resist heteroskedastic ε_{it}'s. Their scalings could be modified in the same spirit as in the Wooldridge or Häggström and Laitila test statistics. Such robustification typically would not affect their asymptotic behavior, hence their optimality properties in case the actual ε_{it}'s are homoskedastic. In order to avoid overloading this paper, those heteroskedasticity concerns are left for future research.

Although optimal rank tests are the main objective of the paper, their construction requires a thorough asymptotic study of the testing problem under study, which is of independent interest. The exact optimality properties and local power of a generalized version of the Honda test (where residuals are not necessarily the OLS ones), as well as its validity- and efficiency-robustness against violations of Gaussian assumptions (strengthening earlier results by Orme and Yamagata 2006), follow as interesting by-products.

1.2 Outline of the paper.

The paper is organized as follows. In Section 2.1, we collect the main assumptions needed in the sequel. Section 2.2 states the uniform local asymptotic normality (ULAN) result (for fixed density f_1 of the ε_{it}'s) that allows us (Section 3.1) to construct locally and asymptotically optimal parametric tests. Due to the mixture nature of likelihoods under random individual effects, establishing that ULAN result is particularly delicate. The special case of the pseudo-Gaussian tests (optimal under Gaussian densities but valid under finite-variance[1] non-Gaussian

[1]From that point of view, we are reinforcing the result by Orme and Yamagata (2006), which requires finite moments of order $4+\delta$; however, their assumptions on the regressors are more general than ours.

ones), which are shown to be asymptotically equivalent to the Honda test, is investigated in Section 3.2. Those optimal parametric procedures, however, are but an intermediate step in the construction (Sections 4.1 and 4.2) of the more important rank-based optimal tests. Some special cases (van der Warden and Wilcoxon scores) are considered in Section 4.3. Asymptotic relative efficiencies with respect to the pseudo-Gaussian and Honda tests are derived in Section 4.4. Section 5 provides some simulation results assessing the finite-sample performance of the various tests proposed. Finally, the appendix collects the proofs and technical results.

2 Uniform local asymptotic normality.

2.1 Notation and technical assumptions.

Denote by $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,\sigma^2_u;f_1,h_1}$ the probability distribution of the observation $(\mathbf{Y}^{(n)\prime}_1, \mathbf{Y}^{(n)\prime}_2, \ldots, \mathbf{Y}^{(n)\prime}_n)'$, where $\mathbf{Y}^{(n)}_i := (Y_{i1}, \ldots, Y_{iT})'$ is generated by

$$Y_{it} = \mu + \boldsymbol{\beta}'\mathbf{x}_{it} + u_i + \varepsilon_{it} = \boldsymbol{\theta}'\underline{\mathbf{x}}_{it} + u_i + \varepsilon_{it}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T,$$

where $\mathbf{x}_{it} = (x_{1;it}, x_{2;it}, \ldots, x_{K;it})'$ is the K-tuple of explanatory variables for individual i at time t, $\underline{\mathbf{x}}_{it} := (1, \mathbf{x}'_{it})'$, and $\boldsymbol{\theta} := (\mu, \boldsymbol{\beta}')' \in \mathbb{R}^{K+1}$; $\{u_i,\ i = 1, \ldots, n\}$ is an unobservable i.i.d. sequence of individual random effects with mean zero, variance σ^2_u, and density $u \mapsto h(u) := (1/\sigma_u)h_1(u/\sigma_u)$; $\{\varepsilon_{it},\ i = 1, \ldots, n,\ t = 1, \ldots, T\}$ is another unobservable i.i.d. sequence, with density $\varepsilon \mapsto f(\varepsilon) := (1/\sigma)f_1(\varepsilon/\sigma)$, for some scale parameter $\sigma \in \mathbb{R}^+_0$ and f_1 in the class of standardized densities

$$\mathcal{F}_0 := \Big\{ f_1 : \int_{-\infty}^{0} f_1(z)dz = 0.5 = \int_{-1}^{1} f_1(z)dz \Big\}.$$

Under $f_1 \in \mathcal{F}_0$, the ε_{it}'s therefore have median zero and median absolute deviation σ; this standardization which, contrary to the usual one based on the mean and the standard error, avoids all moment assumptions, plays the role of an identification constraint, and has no impact on subsequent results. Finally, the individual effects u_i and the disturbances ε_{jt} are assumed to be mutually independent for all i, j, and t.

Clearly, whenever $\sigma_u = 0$, the probability density h of individual effects has no impact on the distribution of the observation; we emphasize that fact by writing $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,0;f_1}$ or $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};f_1}$ instead of $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};f_1,h_1}$ whenever $\sigma_u = 0$, where the notation $\boldsymbol{\vartheta}$ throughout is used for parameter values of the particular form $(\boldsymbol{\theta}', \sigma^2, 0)'$.

Our derivation of locally asymptotically optimal tests at density f_1 will be based on the uniform local asymptotic normality (ULAN), with respect to $(\boldsymbol{\theta}', \sigma^2, \sigma^2_u)'$, of the families of distributions

$$\mathcal{P}^{(n)}_{f_1,h_1} := \Big\{ \mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,\sigma^2_u;f_1,h_1} : \boldsymbol{\theta} \in \mathbb{R}^{K+1},\ \sigma^2 > 0 \text{ and } \sigma^2_u \geq 0 \Big\}$$

at any $\boldsymbol{\vartheta} = (\boldsymbol{\theta}', \sigma^2, 0)'$. This ULAN property requires some technical assumptions on the innovation density f_1, the asymptotic behavior of the regressors, and the density function h of individual random effects.

ASSUMPTION (A). The density f_1 is such that

(A1) $f_1 \in \mathcal{F}_0$;

(A2) $f_1(z) > 0$ for all $z \in \mathbb{R}$;

(A3) $z \mapsto f_1(z)$ is C^2 on $\mathbb{R}$, with derivatives $\dot{f}_1$ and $\ddot{f}_1$; letting $\phi_{f_1} := -\dot{f}_1/f_1$ and $\psi_{f_1} := \ddot{f}_1/f_1$,

$$\mathcal{I}_\phi(f_1) := \int_{\mathbb{R}} \phi^2_{f_1}(z) f_1(z) dz < \infty, \;\; \mathcal{I}_\psi(f_1) := \int_{\mathbb{R}} \psi^2_{f_1}(z) f_1(z) dz < \infty,$$

$$\mathcal{K}_{\phi\phi} := \int_{\mathbb{R}} z\phi^2_{f_1}(z) f_1(z) dz < \infty, \quad \text{and} \quad \mathcal{J}_\phi(f_1) := \int_{\mathbb{R}} z^2\phi^2_{f_1}(z) f_1(z) dz < \infty.$$

Note that (A3) automatically also entails

$$\mathcal{I}_{\phi\psi}(f_1) := \int_{\mathbb{R}} \psi_{f_1}(z)\phi_{f_1}(z) f_1(z) dz < \infty, \quad \text{and} \quad \mathcal{K}_{\phi\psi}(f_1) := \int_{\mathbb{R}} z\psi_{f_1}(z)\phi_{f_1}(z) f_1(z) dz < \infty.$$

The set of all densities satisfying Assumption (A) will be denoted as $\mathcal{F}_A$. It should be stressed that none of these assumptions requires the existence of any moment for the density f_1. They are satisfied, for instance, for all Student distributions with $\nu > 0$ degrees of freedom, with standardized densities

$$f_1(z) = f_{1,t_\nu}(z) := \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \sqrt{a_\nu/\pi\nu}\,(1 + a_\nu z^2/\nu)^{-(1+\nu)/2},$$

for which

$$\mathcal{I}_\phi(f_{1,t_\nu}) = a_\nu \frac{(\nu+1)}{(\nu+3)}, \quad \mathcal{I}_\psi(f_{1,t_\nu}) = 2a^2_\nu \left(\frac{\nu^4 + 8\nu^3 + 27\nu^2 + 40\nu + 20}{\nu^4 + 15\nu^3 + 71\nu^2 + 105\nu} \right),$$

$$\mathcal{J}_\phi(f_{1,t_\nu}) = 3\,\frac{(\nu+1)}{(\nu+3)}, \quad \mathcal{I}_{\phi\psi}(f_{1,t_\nu}) = 0, \quad \text{and} \quad \mathcal{K}_{\phi\psi}(f_{1,t_\nu}) = 2a_\nu \frac{(\nu+1)(\nu+2)}{(\nu+3)(\nu+5)} \tag{2.1}$$

(the normalizing constant $a_\nu > 0$ is such that $f_1 \in \mathcal{F}_0$; for $\nu > 2$, it is equal to the ratio of the variance of a Student with ν degrees of freedom and that of a rescaled Student with median of squares equal to one). Similarly, for the logistic distribution, with standardized density

$$f_1(z) = \ell_1(z) := \sqrt{b}\, \exp(-\sqrt{b}z)/(1 + \exp(-\sqrt{b}z))^2,$$

$$\mathcal{I}_\phi(\ell_1) = b/3, \;\; \mathcal{I}_\psi(\ell_1) = b^2/5, \;\; \mathcal{J}_\phi(\ell_1) = (12 + \pi^2)/9, \;\; \mathcal{I}_{\phi\psi}(\ell_1) = 0, \;\; \text{and} \;\; \mathcal{K}_{\phi\psi}(\ell_1) = b/2$$

($b = (\ln 3)^2 \approx 1.207$ is such that $\ell_1 \in \mathcal{F}_0$, with variance $\pi^2/3b$). The corresponding values for the Gaussian distribution, with standardized density $f_1(z) = \phi_1(z) := \sqrt{a/2\pi}\, \exp(-az^2/2)$ ($a \approx 0.4549$ such that $\phi_1 \in \mathcal{F}_0$, with variance $1/a$), are

$$\mathcal{I}_\phi(\phi_1) = a, \;\; \mathcal{I}_\psi(\phi_1) = 2a^2, \;\; \mathcal{J}_\phi(\phi_1) = 3, \;\; \mathcal{I}_{\phi\psi}(\phi_1) = 0, \;\; \text{and} \;\; \mathcal{K}_{\phi\psi}(\phi_1) = 2a, \tag{2.2}$$

and can be obtained by taking limits as $\nu \to \infty$ of the Student information quantities in (2.1); similarly, $a_\nu \to a$.

We also need some assumptions on the asymptotic behavior of regression constants. Let $\mathbf{C}^{(n)} := \frac{1}{n}\sum_{i=1}^n \sum_{t=1}^T \mathbf{x}_{it}\mathbf{x}'_{it}$. Denoting by $\mathbf{D}^{(n)}$ the diagonal matrix with diagonal elements $(\mathbf{C}^{(n)})_{11}, \ldots, (\mathbf{C}^{(n)})_{KK}$, define $\mathbf{R}^{(n)} := (\mathbf{D}^{(n)})^{-1/2}\mathbf{C}^{(n)}(\mathbf{D}^{(n)})^{-1/2}$. The following assumption is standard in the context of rank-based inference.

Assumption (B). (B1) The limit $\lim_{n\to\infty} \mathbf{R}^{(n)} =: \mathbf{R}$ exists, is positive definite, and therefore factorizes into $\mathbf{R} = (\mathbf{K}\mathbf{K}')^{-1}$ for some full-rank $K \times K$ matrix $\mathbf{K}$.

(B2) The classical Noether conditions hold: letting $\overline{x}_k^{(n)} := \frac{1}{nT}\sum_{i,t}(\mathbf{x}_{it})_k$, assume, without loss of generality, that $\overline{x}_k^{(n)} = 0$ for all k and n, and that

$$\lim_{n\to\infty} \max\nolimits_{1\le i\le n}(\mathbf{x}_{it})_k^2 / \sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbf{x}_{it})_k^2 = 0, \qquad k=1,\ldots,K,\ \ t=1,\ldots,T.$$

Letting $\mathbf{K}^{(n)} := (\mathbf{D}^{(n)})^{-1/2}\mathbf{K}$, note that $\mathbf{K}^{(n)}$ is also of full rank.

Finally, Assumption (C) is about the (standardized) density h_1 of individual effects. Define $K_{\mathbf{z}}(u,y) := \prod_{t=1}^{T} f_1(z_t - yu)$ and, for $y>0$, $\ddot{K}_{\mathbf{z}}(u,y) := \partial^2 K_{\mathbf{z}}(u,y)/\partial y^2$. Writing $\mathbf{z}$ for $(z_1,\ldots,z_T)' \in \mathbb{R}$, the Fisher information associated with σ_u is

$$\mathcal{I}_{\phi\psi}(f_1,y) := \begin{cases} \dfrac{1}{y^2}\displaystyle\int_{\mathbb{R}^T} \frac{\left[\int_{w=0}^{y}\int_{\mathbb{R}} \ddot{K}_{\mathbf{z}}(u,w)h_1(u)du dw\right]^2}{\int_{\mathbb{R}}\prod_{t=1}^{T} f_1(z_t - yu)h_1(u)du}\, d\mathbf{z} & \text{if } y>0 \\[2ex] T\mathcal{I}_{\psi}(f_1) + 2T(T-1)\mathcal{I}_{\phi}^2(f_1) & \text{if } y=0. \end{cases} \tag{2.3}$$

Assumption (C). (C1) $\int_{\mathbb{R}} u\, h_1(u)du = 0$ and $\int_{\mathbb{R}} u^2 h_1(u)du = 1$;

(C2) $y \mapsto \mathcal{I}_{\phi\psi}(f_1,y)$ (the Fisher information for σ_u) is continuous from the right at $y=0$.

Assumption (C2) actually is an assumption involving the couple (f_1,h_1). For all $f_1 \in \mathcal{F}_A$, let

$$\mathcal{F}_{C|f_1} := \Big\{ h_1 \,|\, h_1 \text{ and } (f_1,h_1) \text{ satisfy Assumptions (C1) and (C2), respectively} \Big\}.$$

2.2 ULAN.

In this section we establish the *uniform local asymptotic normality* (ULAN) result (with respect to intercept, regression coefficient, scale parameter σ^2, and the parameter of interest σ_u^2, for fixed density f_1) on which optimality will be based. For background material on asymptotics, the reader is referred to van der Vaart (1988) (Chapters 6-9) or Le Cam and Yang (2000).

Letting $\boldsymbol{\vartheta} := (\boldsymbol{\theta}', \sigma^2, 0)'$, consider sequences of *local alternatives* of the form $\boldsymbol{\vartheta} + n^{-1/2}\boldsymbol{\nu}^{(n)}\boldsymbol{\tau}^{(n)}$, where

$$\boldsymbol{\nu}^{(n)} := \begin{pmatrix} \boldsymbol{\nu}_1^{(n)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 & 0 \\ \mathbf{0} & 0 & 1 \end{pmatrix}, \ \text{ with } \ \boldsymbol{\nu}_1^{(n)} := \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{K}^{(n)} \end{pmatrix}, \tag{2.4}$$

and $\boldsymbol{\tau}^{(n)} := (\tau_1^{(n)}, \boldsymbol{\tau}_2^{(n)\prime}, \tau_3^{(n)}, \tau_4^{(n)})' \in \mathbb{R}^{K+2}\times\mathbb{R}^+$ (indeed, $\tau_4^{(n)}$ is intrinsically nonnegative) is such that $\sup_n \boldsymbol{\tau}^{(n)\prime}\boldsymbol{\tau}^{(n)} < \infty$; indices $1,\ldots,4$, here and in the sequel, refer to the four subparameters μ, $\boldsymbol{\beta}$, σ^2 and σ_u^2, of which $n^{-1/2}\tau_1^{(n)}$, $n^{-1/2}\mathbf{K}^{(n)}\boldsymbol{\tau}_2^{(n)}$, $n^{-1/2}\tau_3^{(n)}$ and $n^{-1/2}\tau_4^{(n)}$ are local perturbations. Defining the standardized residuals

$$Z_{it} = Z_{it}(\boldsymbol{\theta},\sigma^2) := \sigma^{-1}(Y_{it} - \mu - \boldsymbol{\beta}'\mathbf{x}_{it}), \quad i=1,\ldots,n,\ t=1,\ldots,T,$$

(note that, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,0;f_1}$, $\sigma Z_{it}(\boldsymbol{\theta},\sigma^2)$ coincides with ε_{it}), let

$$\boldsymbol{\Delta}^{(n)}_{f_1}(\boldsymbol{\vartheta}) := \begin{pmatrix} \Delta^{(n)}_{f_1;1}(\boldsymbol{\vartheta}) \\ \boldsymbol{\Delta}^{(n)}_{f_1;2}(\boldsymbol{\vartheta}) \\ \Delta^{(n)}_{f_1;3}(\boldsymbol{\vartheta}) \\ \Delta^{(n)}_{f_1;4}(\boldsymbol{\vartheta}) \end{pmatrix} := \begin{pmatrix} \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n}\sum_{t=1}^{T} \phi_{f_1}(Z_{it}) \\ \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n}\sum_{t=1}^{T} \phi_{f_1}(Z_{it})(\mathbf{K}^{(n)})'\mathbf{x}_{it} \\ \frac{1}{2\sigma^2\sqrt{n}} \sum_{i=1}^{n}\sum_{t=1}^{T} (Z_{it}\phi_{f_1}(Z_{it})-1) \\ \frac{1}{2\sigma^2\sqrt{n}} \sum_{i=1}^{n} \left\{ \sum_{t=1}^{T} \psi_{f_1}(Z_{it}) + \sum_{t=1}^{T}\sum_{1=l\neq t}^{T} \phi_{f_1}(Z_{it})\phi_{f_1}(Z_{il}) \right\} \end{pmatrix} \tag{2.5}$$

(in a more traditional perspective, and under stronger regularity assumptions, (2.5) can be obtained as the log-likelihood gradient or Rao score vector) and

$$\boldsymbol{\Gamma}_{f_1}(\boldsymbol{\vartheta}) := \begin{pmatrix} \Gamma_{f_1;11}(\boldsymbol{\vartheta}) & \mathbf{0} & \Gamma_{f_1;13}(\boldsymbol{\vartheta}) & \Gamma_{f_1;14}(\boldsymbol{\vartheta}) \\ \mathbf{0} & \boldsymbol{\Gamma}_{f_1;22}(\boldsymbol{\vartheta}) & \mathbf{0} & \mathbf{0} \\ \Gamma_{f_1;13}(\boldsymbol{\vartheta}) & \mathbf{0} & \Gamma_{f_1;33}(\boldsymbol{\vartheta}) & \Gamma_{f_1;34}(\boldsymbol{\vartheta}) \\ \Gamma_{f_1;14}(\boldsymbol{\vartheta}) & \mathbf{0} & \Gamma_{f_1;34}(\boldsymbol{\vartheta}) & \Gamma_{f_1;44}(\boldsymbol{\vartheta}) \end{pmatrix}, \tag{2.6}$$

with

$$\Gamma_{f_1;11}(\boldsymbol{\vartheta}) := \frac{T}{\sigma^2}\mathcal{I}_\phi(f_1), \quad \boldsymbol{\Gamma}_{f_1;22}(\boldsymbol{\vartheta}) := \frac{1}{\sigma^2}\mathcal{I}_\phi(f_1)\boldsymbol{I}_K, \quad \Gamma_{f_1;33}(\boldsymbol{\vartheta}) := \frac{T}{4\sigma^4}(\mathcal{J}_\phi(f_1)-1),$$

$$\Gamma_{f_1;13}(\boldsymbol{\vartheta}) := \frac{T}{2\sigma^3}\mathcal{K}_{\phi\phi}(f_1), \quad \Gamma_{f_1;14}(\boldsymbol{\vartheta}) := \frac{T}{2\sigma^3}\mathcal{I}_{\phi\psi}(f_1), \quad \Gamma_{f_1;34}(\boldsymbol{\vartheta}) := \frac{T}{4\sigma^4}\mathcal{K}_{\phi\psi}(f_1)$$

and

$$\Gamma_{f_1;44}(\boldsymbol{\vartheta}) := \frac{T}{4\sigma^4}\left[\mathcal{I}_\psi(f_1) + 2(T-1)\mathcal{I}^2_\phi(f_1)\right].$$

We have the following result (see the Appendix for a proof).

Proposition 2.1 *Let Assumptions (B) and (C) hold. Fix $f_1 \in \mathcal{F}_A$ and $h_1 \in \mathcal{F}_{C|f_1}$. Then, the family $\mathcal{P}^{(n)}_{f_1,h_1}$ is ULAN (for $n\to\infty$ with fixed T) at any $\boldsymbol{\vartheta} = (\boldsymbol{\theta}',\sigma^2,0)'$, with central sequence $\boldsymbol{\Delta}^{(n)}_{f_1}(\boldsymbol{\vartheta})$ and information matrix $\boldsymbol{\Gamma}_{f_1}(\boldsymbol{\vartheta})$. More precisely, for any sequence $\boldsymbol{\vartheta}^{(n)} = (\mu^{(n)},\boldsymbol{\beta}^{(n)\prime},\sigma^{2(n)},0)'$ such that $\mu^{(n)}-\mu$, $(\mathbf{K}^{(n)})^{-1}(\boldsymbol{\beta}^{(n)}-\boldsymbol{\beta})$ and $\sigma^{2(n)}-\sigma^2$ are $O(n^{-1/2})$, and any bounded sequence $\boldsymbol{\tau}^{(n)} \in \mathbb{R}^{K+2}\times\mathbb{R}^+_0$, we have*

$$\begin{aligned} \Lambda^{(n)}_{\boldsymbol{\vartheta}^{(n)}+n^{-1/2}\boldsymbol{\nu}^{(n)}\boldsymbol{\tau}^{(n)}/\boldsymbol{\vartheta}^{(n)};f_1,h_1} &:= \log\left(d\mathrm{P}^{(n)}_{\boldsymbol{\vartheta}^{(n)}+n^{-1/2}\boldsymbol{\nu}^{(n)}\boldsymbol{\tau}^{(n)};f_1,h_1}/d\mathrm{P}^{(n)}_{\boldsymbol{\vartheta}^{(n)};f_1}\right) \\ &= \boldsymbol{\tau}^{(n)\prime}\boldsymbol{\Delta}^{(n)}_{f_1}(\boldsymbol{\vartheta}^{(n)}) - \frac{1}{2}\boldsymbol{\tau}^{(n)\prime}\boldsymbol{\Gamma}_{f_1}(\boldsymbol{\vartheta})\boldsymbol{\tau}^{(n)} + o_{\mathrm{P}}(1) \end{aligned}$$

and

$$\boldsymbol{\Delta}^{(n)}_{f_1}(\boldsymbol{\vartheta}^{(n)}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0},\boldsymbol{\Gamma}_{f_1}(\boldsymbol{\vartheta})), \qquad \boldsymbol{\vartheta} = (\mu,\boldsymbol{\beta}',\sigma^2,0)',$$

under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta}^{(n)};f_1}$, as $n\to\infty$ with fixed T.

The form of the score for σ_u^2 in (2.5) is essentially the same as that obtained, in a more classical context and under less general assumptions, by Chesher (1984). Note the non-diagonal form of the information matrix $\mathbf{\Gamma}_{f_1}(\boldsymbol{\vartheta})$ in (2.6), which implies that the intercept parameter μ and the scale parameters σ^2 and σ_u^2 are not mutually information-orthogonal. The notation $\mathbf{\Delta}_{f_1}^{(n)}(\boldsymbol{\vartheta})$ and $\mathbf{\Gamma}_{f_1}(\boldsymbol{\vartheta})$ emphasizes the fact that the central sequence and the information matrix do not depend on the density h_1 of individual effects. It may look strange, at first sight, that the information matrix $\mathbf{\Gamma}_{f_1;22}(\boldsymbol{\vartheta})$ for $\boldsymbol{\beta}$ does not involve the regression constants $\mathbf{x}_{it}$. The reason for this somewhat counterintuitive fact is that the regression constants are taken into account, via the $\mathbf{K}^{(n)}$ matrices, in the definition of the local perturbations of $\boldsymbol{\beta}$; see (2.4). This considerably simplifies the notation.

3 Optimal parametric and pseudo-Gaussian tests.

We are interested in testing the null hypothesis of no individual effects ($\sigma_u^2 = 0$), with unspecified standardized error density in $\mathcal{F}_0$, unspecified $\boldsymbol{\theta}$, and unspecified error scale σ—more formally, a sequence, indexed by n, of collections of distributions of the form

$$\mathcal{H}_0^{(n)} := \bigcup_{g_1 \in \mathcal{F}_0} \mathcal{H}_{g_1}^{(n)} := \bigcup_{g_1 \in \mathcal{F}_0} \bigcup_{\boldsymbol{\theta} \in \mathbb{R}^{K+1}} \bigcup_{\sigma^2 > 0} \{\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,0;g_1}\},$$

optimality being sought against alternatives of the form (individual effects with unspecified density h, unspecified $\boldsymbol{\theta}$ and σ, specified standardized density f_1 of the noise)

$$\bigcup_{\boldsymbol{\theta} \in \mathbb{R}^{K+1}} \bigcup_{\sigma^2 > 0} \bigcup_{\sigma_u^2 > 0} \bigcup_{h_1 \in \mathcal{F}_{C|f_1}} \{\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,\sigma_u^2;f_1,h_1}\}$$

for some chosen "reference" density $f_1 \in \mathcal{F}_A$. The parameters $\mu, \boldsymbol{\beta}$, and σ^2 thus are nuisance parameters, while σ_u^2 is the parameter of interest. The asymptotic covariances, in (2.6), between the μ-part $\Delta_{f_1;1}^{(n)}(\boldsymbol{\vartheta})$, the σ^2-part $\Delta_{f_1;3}^{(n)}(\boldsymbol{\vartheta})$ and the σ_u^2-part $\Delta_{f_1;4}^{(n)}(\boldsymbol{\vartheta})$ of (2.5) are not zero. This implies, via Le Cam's Third Lemma, that a local perturbation of μ or σ^2 has the same asymptotic impact on $\Delta_{f_1;4}^{(n)}(\boldsymbol{\vartheta}_0)$ as a local perturbation of σ_u^2: hence, the cost of not knowing the actual values of μ or σ^2 when performing inference on σ_u^2, in general (the Gaussian case, as far as μ is concerned, is an exception), will be strictly positive.

3.1 Optimal parametric tests.

ULAN and the convergence of local experiments to the trivariate Gaussian shifts

$$\mathbf{\Delta} := (\Delta_1, \Delta_3, \Delta_4)' \sim \mathcal{N}\left(\mathbf{\Gamma}(\boldsymbol{\vartheta})(\tau_1, \tau_3, \tau_4)', \mathbf{\Gamma}(\boldsymbol{\vartheta})\right), \quad (\tau_1, \tau_3, \tau_4)' \in \mathbb{R}^2 \times \mathbb{R}^+, \tag{3.7}$$

with $\mathbf{\Gamma}(\boldsymbol{\vartheta}) = (\Gamma_{ij}(\boldsymbol{\vartheta}))$, $i, j = 1, 3, 4$. imply that locally optimal inference on σ_u^2, in the presence of unspecified μ and σ^2, should be based on the residual of the regression, in (3.7), of Δ_4 with respect to $(\Delta_1, \Delta_3)'$, computed at $\Delta_{f_1;4}^{(n)}(\boldsymbol{\vartheta})$ and $(\Delta_{f_1;1}^{(n)}(\boldsymbol{\vartheta}), \Delta_{f_1;3}^{(n)}(\boldsymbol{\vartheta}))'$ (see, for instance, Le Cam (1986), Chapter 11). That residual, called the *efficient central sequence* for σ_u^2, takes the form

$$\begin{aligned}
\Delta_{f_1;4}^{*(n)}(\boldsymbol{\vartheta}) &:= \Delta_{f_1;4}^{(n)}(\boldsymbol{\vartheta}) - (\Gamma_{f_1;14}(\boldsymbol{\vartheta}), \Gamma_{f_1;34}(\boldsymbol{\vartheta})) \begin{pmatrix} \Gamma_{f_1;11}(\boldsymbol{\vartheta}) & \Gamma_{f_1;13}(\boldsymbol{\vartheta}) \\ \Gamma_{f_1;13}(\boldsymbol{\vartheta}) & \Gamma_{f_1;33}(\boldsymbol{\vartheta}) \end{pmatrix}^{-1} \begin{pmatrix} \Delta_{f_1;1}^{(n)}(\boldsymbol{\vartheta}) \\ \Delta_{f_1;3}^{(n)}(\boldsymbol{\vartheta}) \end{pmatrix} \\
&= \Delta_{f_1;4}^{(n)}(\boldsymbol{\vartheta}) - \Gamma_{f_1;1}^{*}(\boldsymbol{\vartheta})\Delta_{f_1;1}^{(n)}(\boldsymbol{\vartheta}) - \Gamma_{f_1;3}^{*}(\boldsymbol{\vartheta})\Delta_{f_1;3}^{(n)}(\boldsymbol{\vartheta}) \qquad (3.8)
\end{aligned}$$

with

$$\Gamma^*_{f_1;1}(\boldsymbol{\vartheta}) := \frac{\Gamma_{f_1;14}(\boldsymbol{\vartheta})\Gamma_{f_1;33}(\boldsymbol{\vartheta}) - \Gamma_{f_1;34}(\boldsymbol{\vartheta})\Gamma_{f_1;13}(\boldsymbol{\vartheta})}{\Gamma_{f_1;11}(\boldsymbol{\vartheta})\Gamma_{f_1;33}(\boldsymbol{\vartheta}) - \Gamma^2_{f_1;13}(\boldsymbol{\vartheta})}$$

and

$$\Gamma^*_{f_1;3}(\boldsymbol{\vartheta}) := \frac{\Gamma_{f_1;11}(\boldsymbol{\vartheta})\Gamma_{f_1;34}(\boldsymbol{\vartheta}) - \Gamma_{f_1;14}(\boldsymbol{\vartheta})\Gamma_{f_1;13}(\boldsymbol{\vartheta})}{\Gamma_{f_1;11}(\boldsymbol{\vartheta})\Gamma_{f_1;33}(\boldsymbol{\vartheta}) - \Gamma^2_{f_1;13}(\boldsymbol{\vartheta})}.$$

Under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};f_1}$, $\Delta^{*(n)}_{f_1;4}(\boldsymbol{\vartheta})$ is asymptotically normal, with mean zero and variance[2]

$$\begin{aligned}\Gamma^*_{f_1;44}(\boldsymbol{\vartheta}) &:= \Gamma_{f_1;44}(\boldsymbol{\vartheta}) - (\Gamma_{f_1;14}(\boldsymbol{\vartheta}), \Gamma_{f_1;34}(\boldsymbol{\vartheta})) \begin{pmatrix} \Gamma_{f_1;11}(\boldsymbol{\vartheta}) & \Gamma_{f_1;13}(\boldsymbol{\vartheta}) \\ \Gamma_{f_1;13}(\boldsymbol{\vartheta}) & \Gamma_{f_1;33}(\boldsymbol{\vartheta}) \end{pmatrix}^{-1} \begin{pmatrix} \Gamma_{f_1;14}(\boldsymbol{\vartheta}) \\ \Gamma_{f_1;34}(\boldsymbol{\vartheta}) \end{pmatrix} \\ &= \Gamma_{f_1;44}(\boldsymbol{\vartheta}) - \Gamma^*_{f_1;1}(\boldsymbol{\vartheta})\Gamma_{f_1;14}(\boldsymbol{\vartheta}) - \Gamma^*_{f_1;3}(\boldsymbol{\vartheta})\Gamma_{f_1;34}(\boldsymbol{\vartheta}). \end{aligned} \tag{3.9}$$

Next, denoting by $\boldsymbol{\upsilon}^{(n)}$ a sequence of full-rank $k \times k$ matrices such that $\|(\boldsymbol{\upsilon}^{(n)})^{-1}\| \to 0$ as $n \to \infty$,[3] recall that a sequence $\hat{\boldsymbol{\gamma}}^{(n)}$ of estimators, defined over a sequence of experiments $\{\mathrm{P}^{(n)}_{\boldsymbol{\gamma}} \,|\, \boldsymbol{\gamma} \in \boldsymbol{\Gamma}\}$ with k-dimensional parameter $\boldsymbol{\gamma}$, is called $\boldsymbol{\upsilon}^{(n)}$-consistent (for $\boldsymbol{\gamma}$) and *asymptotically discrete* if, under $\mathrm{P}^{(n)}_{\boldsymbol{\gamma}}$, as $n \to \infty$,

(D1) $\boldsymbol{\upsilon}^{(n)}(\hat{\boldsymbol{\gamma}}^{(n)} - \boldsymbol{\gamma}) = O_{\mathrm{P}}(1)$,

(D2) the number of possible values of $\hat{\boldsymbol{\gamma}}^{(n)}$ in any $O((\boldsymbol{\upsilon}^{(n)})^{-1})$ sequence of balls[4] centered at $\boldsymbol{\gamma}$ is bounded as $n \to \infty$.

Any $\boldsymbol{\upsilon}^{(n)}$-consistent estimator $\hat{\boldsymbol{\gamma}}^{(n)}$ is easily turned into an asymptotically discrete one $\hat{\boldsymbol{\gamma}}^{(n)}_{\#}$ by defining, for some arbitrary positive discretization constant c that does not depend on n, the k coordinates of $\boldsymbol{\upsilon}^{(n)}\hat{\boldsymbol{\gamma}}^{(n)}_{\#}$ as $(\boldsymbol{\upsilon}^{(n)}\hat{\boldsymbol{\gamma}}^{(n)}_{\#})_j := c^{-1}\,\mathrm{sgn}((\boldsymbol{\upsilon}^{(n)}\hat{\boldsymbol{\gamma}}^{(n)})_j)\lceil c|(\boldsymbol{\upsilon}^{(n)}\hat{\boldsymbol{\gamma}}^{(n)})_j|\rceil$, $j = 1, \ldots, k$, where $\lceil z \rceil$ stands for the smallest integer larger than or equal to z. The discretized estimator $\hat{\boldsymbol{\gamma}}^{(n)}_{\#}$ follows from transforming that point back via $(\boldsymbol{\upsilon}^{(n)})^{-1}$. Subscripts $_\#$ in the sequel indicate such discretization. However, (D2) is a purely theoretical requirement, which is needed in asymptotic statements but has no practical implications under fixed n, as the discretization constant c can be chosen arbitrarily large. In this paper, we have $\boldsymbol{\upsilon}^{(n)} = n^{1/2}(\boldsymbol{\nu}^{(n)})^{-1}$.

Classical results on ULAN families (see, e.g., Chapter 11 of Le Cam 1986) then show that, for any fixed $f_1 \in \mathcal{F}_A$, locally uniformly asymptotically optimal (*most stringent*) tests of $\mathcal{H}^{(n)}_{f_1} = \bigcup_{\boldsymbol{\theta}} \bigcup_{\sigma^2} \{\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,0;f_1}\}$ can be based on $\Delta^{*(n)}_{f_1;4}(\hat{\boldsymbol{\vartheta}}_{\#})$, hence on $T^{*(n)}_{f_1}(\hat{\boldsymbol{\vartheta}}_{\#})$, with

$$T^{*(n)}_{f_1}(\boldsymbol{\vartheta}) := \left(\Gamma^*_{f_1;44}(\boldsymbol{\vartheta})\right)^{-1/2}\Delta^{*(n)}_{f_1;4}(\boldsymbol{\vartheta}) \tag{3.10}$$

and $\hat{\boldsymbol{\vartheta}}_{\#} := (\hat{\boldsymbol{\theta}}^{(n)\prime}_{\#}, \hat{\sigma}^{2(n)}_{\#}, 0)'$ where $\hat{\boldsymbol{\theta}}^{(n)}_{\#}$ and $\hat{\sigma}^{2(n)}_{\#}$ are estimators satisfying (D1) and (D2) under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};f_1}$, $\boldsymbol{\vartheta} = (\boldsymbol{\theta}', \sigma^2, 0)'$. More precisely, we have the following result.

Proposition 3.1 *Let Assumptions (B) and (C) hold, assume that $\hat{\boldsymbol{\theta}}_{\#}$ satisfies (D1) and (D2), and fix $f_1 \in \mathcal{F}_A$. Then,*

[2]Recall that the notation $\boldsymbol{\vartheta}$ is used for null values $(\boldsymbol{\theta}', \sigma^2, 0)'$ of the parameter only.

[3]Any traditional norm can be considered here, such as the Frobenius or spectral ones.

[4]That is, a sequence of balls $B^{(n)}$ such that $\boldsymbol{\upsilon}^{(n)}B^{(n)}$ is bounded uniformly in n.

(i) for any $\boldsymbol{\vartheta} = (\boldsymbol{\theta}', \sigma^2, 0)'$, $T^{*(n)}_{f_1}(\hat{\boldsymbol{\vartheta}}_\#) = T^{*(n)}_{f_1}(\boldsymbol{\vartheta}) + o_{\rm P}(1)$ *is asymptotically normal, with mean zero and mean* $\left(\Gamma^*_{f_1;44}(\boldsymbol{\vartheta})\right)^{1/2}\tau_4$ *under* ${\rm P}^{(n)}_{\boldsymbol{\vartheta};f_1}$ *and* ${\rm P}^{(n)}_{\boldsymbol{\vartheta}+n^{-1/2}\boldsymbol{\nu}^{(n)}\boldsymbol{\tau};f_1,h_1}$ *($h_1 \in \mathcal{F}_{C|f_1}$), respectively, and variance one under both;*

(ii) the sequence of tests $\phi^{(n)}_{f_1}$ *rejecting the null hypothesis* $\mathcal{H}^{(n)}_{f_1}$ *as soon as* $T^{*(n)}_{f_1}(\hat{\boldsymbol{\vartheta}}_\#)$ *exceeds the* $(1-\alpha)$*-quantile of the standard normal distribution, is locally asymptotically most stringent, at asymptotic probability level* α*, for* $\mathcal{H}^{(n)}_{f_1}$ *against alternatives of the form* $\bigcup_{\boldsymbol{\theta}} \bigcup_{\sigma^2} \bigcup_{\sigma^2_u > 0} \bigcup_{h_1 \in \mathcal{F}_{C|f_1}} \left\{{\rm P}^{(n)}_{\boldsymbol{\theta},\sigma^2,\sigma^2_u;f_1,h_1}\right\}$.

An important advantage of the proposed test statistic $T^{*(n)}_{f_1}(\hat{\boldsymbol{\vartheta}}_\#)$ is that it does not require specifying the (standardized) density h_1 of random effects; on the other hand, its form, which involves several score functions (ϕ_{f_1}, ψ_{f_1}), is rather complicated. The Gaussian case is an exception, however, as $\phi_{f_1}(z)$ and $\psi_{f_1}(z)$ reduce, for $f_1 = \phi_1$, to az and $a(az^2-1)$, respectively. One easily checks that the Gaussian versions of $\Delta^{(n)}_{f_1;4}$, (3.8), (3.9) and (3.10) are

$$\Delta^{(n)}_{\phi_1;4}(\boldsymbol{\vartheta}) = \frac{a^2}{2\sigma^2\sqrt{n}} \sum_{i=1}^{n} \Big\{ \sum_{t=1}^{T} (Z^2_{it} - 1/a) + \sum_{t=1}^{T} \sum_{1=l\neq t}^{T} Z_{it} Z_{il} \Big\}, \tag{3.11}$$

$$\Delta^{*(n)}_{\phi_1;4}(\boldsymbol{\vartheta}) = \frac{a^2}{2\sigma^2\sqrt{n}} \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{l\neq t=1}^{T} Z_{it}Z_{il}, \qquad \Gamma^*_{\phi_1,44}(\boldsymbol{\vartheta}) = \frac{a^2 T(T-1)}{2\sigma^4},$$

and hence

$$T^{*(n)}_{\phi_1}(\boldsymbol{\vartheta}) = \frac{a}{\sqrt{2nT(T-1)}} \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{l\neq t=1}^{T} Z_{it}Z_{il}, \tag{3.12}$$

respectively.

3.2 Pseudo-Gaussian tests.

The Gaussian central sequence $\Delta^{*(n)}_{\phi_1;4}(\boldsymbol{\vartheta})$ in (3.11) allows for asymptotically optimal tests under $f_1 = \phi_1$, hence for efficient detection of random individual effects in the parametric Gaussian model characterized by Gaussian disturbances. Relaxing that Gaussian assumption, and extending the validity of the Gaussian optimal test to densities g_1 in a broad class of densities is of course highly desirable. Let us show that this is indeed possible, and that a slight[5] modification, $\Delta^{\dagger(n)}_{\phi_1;4}$, say, of the efficient central sequence $\Delta^{*(n)}_{\phi_1;4}$ leads to a *pseudo-Gaussian* test, remaining valid when the actual density g_1 belongs to the class $\mathcal{F}^2_A$ of all densities in $\mathcal{F}_A$ with finite variance. That result extends the one by Orme and Yamagata (2006), for which finite moments of order $4+\delta$ and OLS estimators are required.[6]

Writing again Z_{it} for $Z_{it}(\boldsymbol{\theta}, \sigma^2)$, define

$$\Delta^{\dagger(n)}_{\phi_1;4}(\boldsymbol{\vartheta}) := \frac{a^2}{2\sigma^2\sqrt{n}} \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{l\neq t=1}^{T} (Z_{it} - m^{(n)}_1)(Z_{il} - m^{(n)}_1), \tag{3.13}$$

[5]That modification indeed should be $o_{\rm P}(1)$ whenever the actual density is Gaussian, thus yielding another version of the (Gaussian) efficient central sequence.

[6]The assumptions Orme and Yamagata (2006) are imposing on the regression constants are, however, more general than those considered here.

where $m_1^{(n)} = m_1^{(n)}(\boldsymbol{\theta}, \sigma^2) := (nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T Z_{it}(\boldsymbol{\theta}, \sigma^2)$ is a root-n consistent estimator, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g_1}$, of $\mu_1(g_1) := \int_{\mathbb{R}} z g_1(z) dz$ (a finite quantity for any $g_1 \in \mathcal{F}_A^2$). Decomposing $Z_{it} - m_1^{(n)}$ into $(Z_{it} - \mu_1(g_1)) + (\mu_1(g_1) - m_1^{(n)})$, it easily follows from this consistency that

$$\Delta^{\dagger(n)}_{\phi_1;4}(\boldsymbol{\vartheta}) = \frac{a^2}{2\sigma^2\sqrt{n}} \sum_{i=1}^n \sum_{t=1}^T \sum_{l \neq t=1}^T (Z_{it} - \mu_1(g_1))(Z_{il} - \mu_1(g_1)) + o_{\mathrm{P}}(1), \tag{3.14}$$

as $n \to \infty$ under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g_1}$. As a direct consequence, $\Delta^{\dagger(n)}_{\phi_1;4}(\boldsymbol{\vartheta})$, still under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g_1}$, is asymptotically normal with mean zero and variance

$$\Gamma^{\dagger}_{\phi_1;g_1;44}(\boldsymbol{\vartheta}) \quad = \quad \frac{a^4 T(T-1)\sigma^4(g_1)}{2\sigma^4}, \quad \text{with} \quad \sigma^2(g_1) := \int (z - \mu_1(g_1))^2 g_1(z) dz. \tag{3.15}$$

Introducing the non-standardized centered residuals

$$W_{it}(\boldsymbol{\beta}) := \sigma(Z_{it}(\boldsymbol{\theta}, \sigma^2) - m_1^{(n)}) = Y_{it} - \boldsymbol{\beta}'\mathbf{x}_{it} - (nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T Y_{it}$$

(recall from Assumption (B2) that $\sum_{i=1}^n \sum_{t=1}^T \mathbf{x}_{it} = \mathbf{0}$), a pseudo-Gaussian test may then be based on a statistic of the form

$$T^{\dagger(n)}_{\phi_1;g_1}(\boldsymbol{\vartheta}) \quad := \quad (\Gamma^{\dagger}_{\phi_1;g_1;44}(\boldsymbol{\vartheta}))^{-1/2} \Delta^{\dagger(n)}_{\phi_1;4}(\boldsymbol{\vartheta}) \tag{3.16}$$

$$:= \quad \frac{1}{\sigma^2(g)\sqrt{2nT(T-1)}} \sum_{i=1}^n \sum_{t=1}^T \sum_{l \neq t=1}^T W_{it}(\boldsymbol{\beta}) W_{il}(\boldsymbol{\beta}) =: T^{\dagger(n)}_{\mathcal{N};g}(\boldsymbol{\beta}), \tag{3.17}$$

where the density g is defined through $g(z) = \sigma^{-1} g_1(z/\sigma)$; compare (3.16) with the Gaussian version (obtained with $f_1 = \phi_1$) of (3.10), and (3.12) with (3.17). Clearly, $T^{\dagger(n)}_{\mathcal{N};g}(\boldsymbol{\beta})$ does not depend on μ nor on σ, which justifies the notation.

In practice, the pseudo-Gaussian tests will be based on the statistics

$$T^{\dagger(n)}_{\mathcal{N}} := \frac{1}{s^2\sqrt{2nT(T-1)}} \sum_{i=1}^n \sum_{t=1}^T \sum_{l \neq t=1}^T W_{it}(\hat{\boldsymbol{\beta}}_\#) W_{il}(\hat{\boldsymbol{\beta}}_\#),$$

where $\hat{\boldsymbol{\beta}}_\#$ is an arbitrary $n^{1/2}(\mathbf{K}^{(n)})^{-1}$-consistent and locally asymptotically discrete estimator of $\boldsymbol{\beta}$ (see (D1-D2)), and $s^2 := (nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T W_{it}^2(\hat{\boldsymbol{\beta}}_\#)$ is the empirical variance of the $W_{it}(\hat{\boldsymbol{\beta}}_\#)$'s. Showing that the difference $T^{\dagger(n)}_{\mathcal{N}} - T^{\dagger(n)}_{\mathcal{N};g}(\boldsymbol{\beta})$ is $o_{\mathrm{P}}(1)$ as $n \to \infty$ under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g_1}$ mainly requires the asymptotic linearity property of Lemma 6.2, which itself follows from the block-diagonality of the covariance matrix in the asymptotically normal distribution of $(\boldsymbol{\Delta}^{(n)}_{g_1;2}(\boldsymbol{\vartheta}), \Delta^{\dagger(n)}_{\phi_1;4}(\boldsymbol{\vartheta}))'$ under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g_1}$. As for the substitution of s^2 for $\sigma^2(g)$, it is simply controlled by the continuous mapping theorem.

Summing up, we have the following result.

Proposition 3.2 *Let Assumptions (B) and (C) hold and let $\hat{\boldsymbol{\beta}}_\#$ be such that $\hat{\boldsymbol{\beta}}_\#$ satisfies (D1)-(D2) under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g_1}$ for any $g_1 \in \mathcal{F}_A^2$. Then,* *(i) for any* $\boldsymbol{\vartheta} = (\mu, \boldsymbol{\beta}', \sigma^2, 0)'$ *and* $g_1 \in \mathcal{F}_A^2$, $T_{\mathcal{N}}^{\dagger(n)} = T_{\mathcal{N};g}^{\dagger(n)}(\boldsymbol{\beta}) + o_{\rm P}(1)$ *is asymptotically normal, with mean zero under* ${\rm P}^{(n)}_{\mu,\boldsymbol{\beta},\sigma^2,0;g_1}$*, mean* $\sqrt{T(T-1)/2\sigma^4(g)}\,\tau_4$ *under* ${\rm P}^{(n)}_{\boldsymbol{\vartheta}+n^{-1/2}\boldsymbol{\nu}^{(n)}\boldsymbol{\tau};g_1,h_1}$ *(*$h_1 \in \mathcal{F}_{C|g_1}$*), and variance one under both;*

(ii) the sequence of tests $\phi_{\mathcal{N}}^{\dagger(n)}$ *rejecting the null hypothesis* $\mathcal{H}_A^{(n)2} := \bigcup_{g_1\in\mathcal{F}_A^2}\bigcup_{\boldsymbol{\theta}}\bigcup_{\sigma^2}\{{\rm P}^{(n)}_{\boldsymbol{\theta},\sigma^2,0;g_1}\}$ *as soon as* $T_{\mathcal{N}}^{\dagger(n)}$ *exceeds the* $(1-\alpha)$*-quantile of the standard normal distribution is locally asymptotically most stringent, at asymptotic probability level* α*, for* $\mathcal{H}_A^{(n)2}$ *against alternatives of the form* $\bigcup_{\boldsymbol{\theta}}\bigcup_{\sigma^2}\bigcup_{\sigma_u^2>0}\bigcup_{h_1\in\mathcal{F}_{C|\phi_1}}\{{\rm P}^{(n)}_{\boldsymbol{\theta},\sigma^2,\sigma_u^2;\phi_1,h_1}\}$*.*

Since $\sum_{i=1}^n\sum_{t=1}^T \mathbf{x}_{it} = \mathbf{0}$, the OLS estimator of $\boldsymbol{\theta}$ is of the form $\hat{\boldsymbol{\theta}}^{(n)}_{\rm OLS} = (\hat{\mu}^{(n)}_{\rm OLS}, \hat{\boldsymbol{\beta}}^{(n)\prime}_{\rm OLS})' = ((nT)^{-1}\sum_{i=1}^n\sum_{t=1}^T Y_{it}, \hat{\boldsymbol{\beta}}^{(n)\prime}_{\rm OLS})'$, which implies that the OLS residuals $\varepsilon_{it}^{\rm OLS}$ and the $W_{it}(\hat{\boldsymbol{\beta}}^{(n)}_{\rm OLS})$'s coincide. The resulting test statistic $T_{\mathcal{N}}^{\dagger(n)}$ thus is nothing else but the Honda test statistic $T^{(n)}_{\rm Honda}$. For any other estimator $\hat{\boldsymbol{\beta}}^{(n)}$ with discretized version $\hat{\boldsymbol{\beta}}_{\#}$ satisfying Assumptions (D1) and (D2) (such as the LAD estimator), $T_{\mathcal{N}}^{\dagger(n)}$ yields a new pseudo-Gaussian test statistic.

4 Optimal rank tests.

4.1 Rank-based versions of central sequences.

A serious drawback of the parametric tests of Section 3.1 is that their validity is restricted, in general, to the correctly specified density f_1 they have been constructed for. As for the pseudo-Gaussian tests of Section 3.2, they still require finite moments of order two; moreover, for fixed n, the quality of the Gaussian approximation they are based on is likely to depend on the actual underlying density.

Since a correct specification of the actual density g_1 in practice is highly unrealistic, the problem has to be considered from a semiparametric point of view, where g_1 plays the role of a nuisance. A general result by Hallin and Werker (2003) suggests that, in such context, semiparametrically efficient (in the sense of Bickel et al. (1993), at selected $g_1 = f_1$) tests can be obtained by conditioning the f_1-central sequence on the maximal invariant associated with some appropriate generating group.

More precisely, note that the null hypothesis $\mathcal{H}_{\boldsymbol{\beta}}^{(n)} := \bigcup_{g_1\in\mathcal{F}_0}\bigcup_{\mu\in\mathbb{R}}\bigcup_{\sigma^2>0}\{{\rm P}^{(n)}_{\mu,\boldsymbol{\beta},\sigma^2,0;g_1}\}$ is invariant under the action of the group $\mathcal{G}_{\boldsymbol{\beta}}^{(nT)}, \circ$ of all transformations $\mathcal{G}_l$ of $\mathbb{R}^{nT}$ such that

$$\mathcal{G}_l(y_{11},\ldots,y_{nT}) := (\boldsymbol{\beta}'\mathbf{x}_{11} + l(y_{11} - \boldsymbol{\beta}'\mathbf{x}_{11}),\ldots,\boldsymbol{\beta}'\mathbf{x}_{nT} + l(y_{nT} - \boldsymbol{\beta}'\mathbf{x}_{nT})),$$

where $z \mapsto l(z)$ is continuous and monotone increasing and $\lim_{z\to\pm\infty} l(z) = \pm\infty$ (the *order-preserving group*); note that the $\mathcal{G}_{\boldsymbol{\beta}}^{(nT)}, \circ$ is a generating group for $\mathcal{H}_{\boldsymbol{\beta}}^{(n)}$, and has maximal invariant $(R_{11}^{(n)}(\boldsymbol{\beta}),\ldots,R_{nT}^{(n)}(\boldsymbol{\beta}))$, where $R_{it}^{(n)}(\boldsymbol{\beta})$ denotes the rank of the residual $Z_{it}(\boldsymbol{\theta},\sigma^2)$ among $Z_{11}(\boldsymbol{\theta},\sigma^2),\ldots,Z_{nT}(\boldsymbol{\theta},\sigma^2)$ (those ranks are also those of the $(Y_{it} - \mu - \boldsymbol{\beta}'\mathbf{x}_{it})$'s, hence do not depend on μ and σ^2, which justifies the notation). The idea of considering tests that are measurable with respect to those ranks looks quite natural and appealing, since such tests would be distribution-free (or asymptotically so, since the unspecified $\boldsymbol{\beta}$ will have to be replaced with some estimator) and hopefully semiparametrically efficient. As we shall show, this is indeed the case, and semiparametric efficiency moreover turns out to match parametric efficiency at the selected reference density f_1.

The rank-based version of the σ_u^2-efficient central sequence $\Delta^{*(n)}_{f_1;4}(\boldsymbol{\vartheta})$ we plan to base our tests on is (the notation $\underset{\sim}{\Delta}{}^{*(n)}_{f_1;4}(\boldsymbol{\beta}, \sigma^2)$ reflects the fact that it only depends on $\boldsymbol{\beta}$ and σ^2)

$$\underset{\sim}{\Delta}{}^{*(n)}_{f_1;4}(\boldsymbol{\beta}, \sigma^2) \;\; := \;\; \frac{1}{2\sigma^2\sqrt{n}} \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{l \neq t=1}^{T} \Big\{ \varphi_{f_1}\Big(\frac{R_{it}^{(n)}(\boldsymbol{\beta})}{N+1}\Big) \varphi_{f_1}\Big(\frac{R_{il}^{(n)}(\boldsymbol{\beta})}{N+1}\Big) - c_{f_1}^{(n)} \Big\}, \qquad (4.1)$$

with $c_{f_1}^{(n)} := \frac{1}{N(N-1)} \sum_{r=1}^{N} \sum_{s\neq r=1}^{N} \varphi_{f_1}(\frac{r}{N+1})\varphi_{f_1}(\frac{s}{N+1})$ and $\varphi_{f_1} := \phi_{f_1} \circ F_1^{-1}$.

Beyond its role in the derivation of the asymptotic distribution of $\underset{\sim}{\Delta}{}^{*(n)}_{f_1;4}(\boldsymbol{\beta}, \sigma^2)$, the following asymptotic representation result shows that $\underset{\sim}{\Delta}{}^{*(n)}_{f_1;4}(\boldsymbol{\beta}, \sigma^2)$ is indeed another version of the efficient central sequence $\Delta^{*(n)}_{f_1;4}(\boldsymbol{\vartheta})$ (compare (3.8) with $\Delta^{*(n)}_{f_1,f_1;4}(\boldsymbol{\vartheta})$ defined in (4.3) below). Writing $\sum_{1\leq r_1 \neq \ldots \neq r_q \leq N}$ for a sum running over the $(N!/(N-q)!)$ ordered q-tuples of distinct integers in $\{1, \ldots, N\}$, let

$$\begin{aligned} s_{f_1}^{2(n)} := \; & \gamma_1 \sum_{r=1}^{N} \sum_{s\neq r=1}^{N} \varphi_{f_1}^2\Big(\frac{r}{N+1}\Big)\varphi_{f_1}^2\Big(\frac{s}{N+1}\Big) + \gamma_2 \sum_{1\leq r\neq s\neq v\leq N} \varphi_{f_1}^2\Big(\frac{r}{N+1}\Big)\varphi_{f_1}\Big(\frac{s}{N+1}\Big)\varphi_{f_1}\Big(\frac{v}{N+1}\Big) \\ & + \gamma_3 \sum_{1\leq r\neq s\neq v\neq w\leq N} \varphi_{f_1}\Big(\frac{r}{N+1}\Big)\varphi_{f_1}\Big(\frac{s}{N+1}\Big)\varphi_{f_1}\Big(\frac{v}{N+1}\Big)\varphi_{f_1}\Big(\frac{w}{N+1}\Big) - \gamma_4 \Big(c_{f_1}^{(n)}\Big)^2, \qquad (4.2) \end{aligned}$$

where

$$\gamma_1 := \frac{T(T-1)}{2N(N-1)}, \quad \gamma_2 := \frac{7T(T-1)(T-2)}{3N(N-1)(N-2)}, \quad \gamma_3 := \frac{T(T-1)(nT^2-(n+4)T+6)}{4N(N-1)(N-2)(N-3)},$$

and

$$\gamma_4 := \frac{T(T-1)(3nT^2-(3n-16)T-32)}{12}.$$

Define the *cross-information coefficient* $\mathcal{I}_\phi(f_1, g_1)$ as

$$\mathcal{I}_\phi(f_1, g_1) \;\; := \;\; \int_0^1 \phi_{f_1}(F_1^{-1}(u))\phi_{g_1}(G_1^{-1}(u))\, du,$$

and denote by $\underset{\sim}{\mathcal{F}}_A$ the class of all densities $f_1 \in \mathcal{F}_A$ such that ϕ_{f_1} can be expressed as the difference of two monotone increasing functions. We then have for the rank-based $\underset{\sim}{\Delta}{}^{*(n)}_{f_1;4}(\boldsymbol{\beta}, \sigma^2)$ the following asymptotic representation result.

Proposition 4.1 *Fix* $\boldsymbol{\vartheta} = (\boldsymbol{\theta}', \sigma^2, 0)'$ *(with* $\boldsymbol{\theta} \in \mathbb{R}^{K+1}$ *and* $\sigma^2 > 0$*),* $f_1 \in \underset{\sim}{\mathcal{F}}_A$*, and* $g_1 \in \mathcal{F}_0$*. Then, (i) under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g_1}$*, as* $n \to \infty$ *with fixed* T*,*

$$\underset{\sim}{\Delta}{}^{*(n)}_{f_1;4}(\boldsymbol{\beta}, \sigma^2) = \mathrm{E}^{(n)}_{\boldsymbol{\vartheta};g_1}[\Delta^{(n)}_{f_1;4}(\boldsymbol{\vartheta}) | R_{11}^{(n)}(\boldsymbol{\beta}), \ldots, R_{nT}^{(n)}(\boldsymbol{\beta})] + o_{L^2}(1) = \Delta^{*(n)}_{f_1,g_1;4}(\boldsymbol{\vartheta}) + o_{L^2}(1),$$

with (denoting by G_1 *the distribution function associated with* g_1*)*

$$\Delta^{*(n)}_{f_1,g_1;4}(\boldsymbol{\vartheta}) := \frac{1}{2\sigma^2\sqrt{n}} \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{1=l\neq t}^{T} \varphi_{f_1}(G_1(Z_{it}))\varphi_{f_1}(G_1(Z_{il})); \qquad (4.3)$$

(ii) under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g_1}$*,* $\underset{\sim}{\Delta}{}^{*(n)}_{f_1;4}(\boldsymbol{\beta}, \sigma^2)$ *has mean zero and variance* $\Gamma^{*(n)}_{f_1;44}(\sigma^2) := \sigma^{-4} s_{f_1}^{2(n)} = \Gamma^{*}_{f_1;44}(\sigma^2) +$ $o(1)$ *as* $n \to \infty$ *with fixed* T*, where* $\Gamma^{*}_{f_1;44}(\sigma^2) := (T(T-1)/2\sigma^4)\mathcal{I}^2_\phi(f_1)$*;*

(iii) $\Delta^{*(n)}_{f_1,g_1;4}(\boldsymbol{\vartheta})$ *is asymptotically normal with mean zero and mean* $\big(T(T-1)/2\sigma^4\big)\,\mathcal{I}^2_\phi(f_1,g_1)\tau_4$ *under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g_1}$ *and* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta}+n^{-1/2}\boldsymbol{\nu}^{(n)}\boldsymbol{\tau};g_1,h_1}$*, respectively, and variance* $\Gamma^*_{f_1;44}(\sigma^2)$ *under both (the claim under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta}+n^{-1/2}\boldsymbol{\nu}^{(n)}\boldsymbol{\tau};g_1,h_1}$ *further requires* $g_1\in\mathcal{F}_A$ *and* $h_1\in\mathcal{F}_{C|g_1}$ *).*

4.2 Optimal rank tests.

The parameters μ, $\boldsymbol{\beta}$, and σ^2 remain unspecified under the null. Since only $\boldsymbol{\beta}$ has an influence on the ranks, a consistent estimator $\hat{\boldsymbol{\beta}}^{(n)}$ has to be substituted for the actual $\boldsymbol{\beta}$ value, yielding *aligned ranks* $R^{(n)}_{it}(\hat{\boldsymbol{\beta}}^{(n)})$. The effect of this alignement procedure is taken care of (Lemma 6.3(ii)) in a similar way as in Section 3, via an asymptotic linearity result (Lemma 6.3(i)); see the Appendix.

Local asymptotic optimality at density f_1 is achieved by the test based on $\underset{\sim}{T}^{*(n)}_{f_1}(\hat{\boldsymbol{\beta}}_\#)$, where

$$\underset{\sim}{T}^{*(n)}_{f_1}(\boldsymbol{\vartheta}) := \Big(\Gamma^{*(n)}_{f_1;44}(\sigma^2)\Big)^{-1/2}\,\underset{\sim}{\Delta}^{*(n)}_{f_1;4}(\boldsymbol{\vartheta}) \qquad (4.4)$$

$$= \frac{1}{2s^{(n)}_{f_1}\sqrt{n}}\sum_{i=1}^n\sum_{t=1}^T\sum_{l\neq t=1}^T\left\{\varphi_{f_1}\Big(\frac{R^{(n)}_{it}(\boldsymbol{\beta})}{N+1}\Big)\varphi_{f_1}\Big(\frac{R^{(n)}_{il}(\boldsymbol{\beta})}{N+1}\Big)-c^{(n)}_{f_1}\right\} =: \underset{\sim}{T}^{*(n)}_{f_1}(\boldsymbol{\beta})$$

and $\hat{\boldsymbol{\beta}}_\#$ satisfies (D1) and (D2). More precisely, we have the following result.

Proposition 4.2 *Let Assumptions (B) and (C) hold, assume that* $\hat{\boldsymbol{\beta}}_\#$ *satisfies (D1) and (D2), and fix* $f_1\in\underset{\sim}{\mathcal{F}}_A$*. Then,*

(i) for any $g_1\in\mathcal{F}_A$*,* $\underset{\sim}{T}^{*(n)}_{f_1}(\hat{\boldsymbol{\beta}}_\#)$ *is asymptotically normal with mean zero and mean*

$$\Big(T(T-1)\mathcal{I}^2_\phi(f_1,g_1)/2\sigma^4\big(\Gamma^*_{f_1;44}(\sigma^2)\big)^{1/2}\Big)\,\tau_4$$

under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,0;g_1}$ *and* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta}+n^{-1/2}\boldsymbol{\nu}^{(n)}\boldsymbol{\tau};g_1,h_1}$ *(*$h_1\in\mathcal{F}_{C|g_1}$*), respectively, and variance one under both.*

(ii) the sequence of tests $\phi^{*(n)}_{f_1}$ *rejecting the null hypothesis* $\mathcal{H}^{(n)}_A := \bigcup_{g_1\in\mathcal{F}_A}\bigcup_{\boldsymbol{\theta}}\bigcup_{\sigma^2}\Big\{\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,0;g_1}\Big\}$ *as soon as* $\underset{\sim}{T}^{(n)}_{f_1}(\hat{\boldsymbol{\beta}}_\#)$ *exceeds the* $(1-\alpha)$*-quantile of the standard normal distribution is locally asymptotically most stringent, at asymptotic probability level* α*, for* $\mathcal{H}^{(n)}_A$ *against alternatives of the form* $\bigcup_{\boldsymbol{\theta}}\bigcup_{\sigma^2}\bigcup_{\sigma^2_u>0}\bigcup_{h_1\in\mathcal{F}_{C|f_1}}\{\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,\sigma^2_u;f_1,h_1}\}$*.*

4.3 The van der Waerden and Wilcoxon test statistics.

The statistic $\underset{\sim}{T}^{(n)}_{f_1}(\hat{\boldsymbol{\beta}}_\#)$ is providing a general form for the optimal rank tests of the null hypothesis of absence of random effects. Important particular cases are the Wilcoxon (logistic scores) and van der Waerden (normal scores) test statistics, which are optimal at logistic and normal densities, respectively.

The van der Waerden tests use the standard normal reference density $f_1=\phi_1$. One easily obtains that $\phi_{f_1}(F_1^{-1}(u)) = a^{1/2}\Phi^{-1}(u)$, where Φ denotes the standard normal distribution function. The resulting rank test statistic is then

$$\underset{\sim}{T}^{*(n)}_{\mathrm{vdW}}(\hat{\boldsymbol{\beta}}_\#) := \underset{\sim}{T}^{*(n)}_{\phi_1}(\hat{\boldsymbol{\beta}}_\#) = \frac{a}{2s^{(n)}_{\phi_1}\sqrt{n}}\sum_{i=1}^n\sum_{t=1}^T\sum_{1=l\neq t}^T\left\{\Phi^{-1}\Big(\frac{R^{(n)}_{it}(\hat{\boldsymbol{\beta}}_\#)}{N+1}\Big)\Phi^{-1}\Big(\frac{R^{(n)}_{il}(\hat{\boldsymbol{\beta}}_\#)}{N+1}\Big)-c^{(n)}_{\phi_1}\right\},$$

with

$$c_{\phi_1}^{(n)} = [N(N-1)]^{-1} \sum_{r=1}^{N} \sum_{s \neq r=1}^{N} \Phi^{-1}\left(\frac{r}{N+1}\right) \Phi^{-1}\left(\frac{s}{N+1}\right). \tag{4.5}$$

In the Wilcoxon case, $f_1(z) = \ell_1(z) := \sqrt{b}\exp(-\sqrt{b}z)/(1+\exp(-\sqrt{b}z))^2$ is a standardized logistic density. One easily checks that $\phi_{f_1}(F_1^{-1}(u)) = b^{1/2}u$. Therefore, the Wilcoxon test statistic takes the form

$$\underset{\sim}{T}_{\mathrm{W}}^{*(n)}(\hat{\boldsymbol{\beta}}_{\#}) := \underset{\sim}{T}_{\ell_1}^{*(n)}(\hat{\boldsymbol{\beta}}_{\#}) = \frac{b}{2s_{\ell_1}^{(n)}\sqrt{n}} \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{1=l\neq t}^{T}\left(\frac{R_{it}^{(n)}(\hat{\boldsymbol{\beta}}_{\#})}{N+1}\frac{R_{il}^{(n)}(\hat{\boldsymbol{\beta}}_{\#})}{N+1} - c_{\ell_1}^{(n)}\right),$$

with $c_{\ell_1}^{(n)} = [N(N-1)]^{-1}\sum_{r=1}^{N}\sum_{s\neq r=1}^{N}\frac{r}{N+1}\frac{s}{N+1}$. It is worth noting that the scale factors a (for van der Waerden) and b (for Wilcoxon) actually cancel out in the final expression of the test statistics, due to the (exact) standardization by $s_{\phi_1}^{(n)}$ and $s_{\ell_1}^{(n)}$, respectively. This confirms that our choice of the median of absolute deviations as a scale parameter in the definition of $\mathcal{F}_0$ has no impact on the results.

4.4 Asymptotic relative efficiencies (AREs).

Propositions 3.2 and 4.2 allow for computing ARE values for the rank tests based on $\underset{\sim}{T}_{f_1}^{*(n)}(\hat{\boldsymbol{\beta}}_{\#})$ with respect, for instance, to the pseudo-Gaussian tests based on $T_{\mathcal{N}}^{\dagger(n)}$ —equivalently, with respect to the Honda test. These AREs, as usual, are obtained as ratios of the squared shifts under local alternatives.

Proposition 4.3 *Let $f_1 \in \underset{\sim}{\mathcal{F}}_A$. Then, the asymptotic relative efficiencies, under $g_1 \in \mathcal{F}_A^2$, of the rank tests based on $\underset{\sim}{T}_{f_1}^{*(n)}(\hat{\boldsymbol{\beta}}_{\#}^{(n)})$ with respect to the pseudo-Gaussian tests based on $T_{\mathcal{N}}^{\dagger(n)}$ or the Honda test based on $T_{Honda}^{(n)}$ are*

$$\mathrm{ARE}_{g_1}(\underset{\sim}{T}_{f_1}^{*(n)}(\hat{\boldsymbol{\beta}}_{\#}^{(n)})/T_{\mathcal{N}}^{\dagger(n)}) = \mathrm{ARE}_{g_1}(\underset{\sim}{T}_{f_1}^{*(n)}(\hat{\boldsymbol{\beta}}_{\#}^{(n)})/T_{Honda}^{(n)}) = \{\sigma^2(g_1)\mathcal{I}_{\phi}^2(f_1,g_1)/\mathcal{I}_{\phi}(f_1)\}^2. \tag{4.6}$$

Strictly speaking, however, the interpretation of (4.6) as an ARE value only holds for standardized individual effect densities $h_1 \in \mathcal{F}_{C|g_1}$.

Numerical values of the ARE values of Proposition 3.2, under t_3, t_5, t_8, t_{10}, t_{20}, normal, and logistic densities, are displayed in Table 1. These values all are good, particularly so under heavy tails (see the Student density with 3 degrees of freedom). Since the AREs in (4.6) are the squares of those obtained in univariate location problems, the Chernoff-Savage (1958) property extends to the problem considered in this paper, showing that our van der Waerden rank test, from the Pitman point of view, uniformly dominates the classical Honda test:

$$\inf_{g_1} \mathrm{ARE}_{g_1}(\underset{\sim}{T}_{f_1}^{*(n)}(\hat{\boldsymbol{\beta}}_{\#}^{(n)})/T_{\mathrm{Honda}}^{(n)}) = 1,$$

that is, the AREs, with respect to Honda, of the van der Waerden tests are uniformly larger than or equal to one, and equal to one in the Gaussian case only.

		actual density g_1						
		f_{1,t_3}	f_{1,t_5}	f_{1,t_8}	$f_{1,t_{10}}$	$f_{1,t_{20}}$	ϕ_1	ℓ_1
scores	f_{1,t_3}	4.0000	1.4923	1.0498	0.9498	0.7917	0.6718	1.0851
	f_{1,t_5}	3.8202	1.5625	1.1667	1.0783	0.9401	0.8370	1.1857
	f_{1,t_8}	3.5284	1.5319	1.1901	1.1158	1.0029	0.9232	1.1993
	$f_{1,t_{10}}$	3.3960	1.5061	1.1870	1.1187	1.0168	0.9476	1.1933
	$f_{1,t_{20}}$	3.0768	1.4272	1.1597	1.1053	1.0292	0.9851	1.1610
	ϕ_1	2.6873	1.3079	1.0986	1.0601	1.0139	1.0000	1.0966
	ℓ_1	3.6091	1.5405	1.1869	1.1101	0.9936	0.9119	1.2026

Table 1: AREs, under Student (3, 5, 8, 10, and 20 degrees of freedom), normal and logistic densities, of various rank tests (based on Student, van der Waerden, and Wilcoxon scores), with respect to the pseudo-Gaussian and Honda tests.

5 Simulations.

In this section, we conduct a Monte Carlo experiment to investigate the finite-sample behavior of our rank tests under a variety of error distributions. More precisely, we considered the model

$$Y_{it} = \mu + \beta_1(x_{1;it} - \overline{x}_1) + \beta_2(x_{2;it} - \overline{x}_2) + u_i + \varepsilon_{it}, \quad i = 1, \ldots, n = 100, \quad t = 1, \ldots, T = 5, \quad (5.7)$$

where

(a) $\mu = \beta_1 = \beta_2 = 1$;

(b) the $x_{1;it}$'s are i.i.d. uniform over $(0, 1)$ and the $x_{2;it}$'s are generated through $x_{2;i0} = 5 + 10\nu_{i0}$ and $x_{2;it} = 0.1t + 0.5x_{2;i,t-1} + \nu_{it}$ for $t \geq 1$, where the ν_{it}'s are i.i.d. uniform over $(-0.5, 0.5)$ (see Nerlove 1971); $\overline{x}_k := (nT)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} x_{k;it}$, $k = 1, 2$;

(c) the ε_{it}'s are i.i.d. with a symmetric density — Gaussian (ϕ_1), Cauchy (t_1), t with 3 degrees of freedom (t_3), logistic (ℓ_1) — or with an asymmetric density — the skew-normal ($\mathcal{SN}$) or skew-t_3 densities ($\mathcal{S}t_3$), both with skewness parameter value $\xi = 10$[7];

(d) when symmetric disturbances are used, the u_i's are i.i.d. Gaussian with mean zero and standard deviation $\sigma_u = 0$ (null hypothesis), 0.2, 0.3, 0.4, or 0.5 (increasingly severe alternatives). When asymmetric disturbances are used, the non-null values of σ_u considered are 0.05, 0.10, 0.15, and 0.20.

For each combination of σ_u and a distribution for the ε_{it}'s, we generated $M = 2,500$ independent samples from (5.7), with the same values of the regressors $x_{k;it}$ over replications. For each replication, the following six tests were performed at nominal probability level $\alpha = 5\%$: two versions of the pseudo-Gaussian test (the first one is the Honda test, based on the usual OLS regression estimator; the second one is based on the least absolute deviation (LAD) estimator of regression), and four of the proposed rank tests, namely the van der Waerden test based on $\underset{\sim}{T}^{*(n)}_{\text{vdW}}(\hat{\boldsymbol{\beta}}^{(n)})$, the two t_ν-score tests based on $\underset{\sim}{T}^{*(n)}_{f_{1,t_\nu}}(\hat{\boldsymbol{\beta}}^{(n)})$ ($\nu = 1, 3$), and the Wilcoxon test based on $\underset{\sim}{T}^{*(n)}_{\text{W}}(\hat{\boldsymbol{\beta}}^{(n)})$, where $\hat{\boldsymbol{\beta}}^{(n)}$ is the LAD estimator of $\boldsymbol{\beta}$. The resulting rejection frequencies are reported in Table 2.

These simulations show that the pseudo-Gaussian tests, though resisting non-Gaussian densities with finite second-order moments, are collapsing under the heavy-tailed t_1 distribution. In

[7]See, for instance, Azzalini and Capitanio (2003) for a definition of skew-normal and skew-t densities.

sharp contrast with this, the proposed rank-based tests always appear to meet the 5% probability level constraint. Empirical power rankings are essentially consistent with the corresponding ARE values from Table 1. For instance, under Gaussian densities, the powers of the t_ν-score rank tests are increasing with ν as expected, whereas the asymptotic optimality of the same tests under the corresponding Student distribution with ν degrees of freedom is confirmed. It also appears from the skew-normal and skew-t simulations that asymmetry significantly improves the superiority of rank tests over Honda's classical procedure. Also note that Honda is fairly insensitive to the choice of the preliminary estimators (OLS or LAD makes little difference).

Test	g_1			σ_u			g_1			σ_u		
		0	.2	.3	.4	.5		0	.2	.3	.4	.5
Honda-$T^{\dagger(n)}_{\phi_1}$		.0420	.3184	.7732	.9764	.9996		.0492	.0916	.2088	.4084	.6680
LAD-$T^{\dagger(n)}_{\phi_1}$		.0412	.3188	.7688	.9752	.9996		.0512	.0904	.2116	.4024	.6588
vdW	ϕ_1	.0488	.3416	.7820	.9812	.9996	ℓ_1	.0552	.1108	.2428	.4448	.7068
t_3		.0504	.2596	.6528	.9372	.9976		.0544	.1224	.2496	.4196	.6832
t_1		.0572	.1396	.3176	.5668	.8256		.0492	.0936	.1500	.2456	.4184
W		.0500	.3144	.7544	.9708	.9996		.0552	.1164	.2600	.4520	.7260
				σ_u						σ_u		
		0	.2	.3	.4	.5		0	.05	.1	.15	.2
Honda-$T^{\dagger(n)}_{\phi_1}$		.0428	.1144	.2528	.5144	.7544		.0504	.0792	.2000	.5180	.8680
LAD-$T^{\dagger(n)}_{\phi_1}$		.0436	.1116	.2500	.5060	.7468		.0532	.0788	.2008	.5132	.8624
vdW	t_3	.0576	.1896	.4332	.7472	.9408	$\mathcal{SN}$	.0580	.1080	.3384	.7252	.9572
t_3		.0580	.2184	.5248	.8356	.9760		.0532	.0636	.0864	.1552	.3196
t_1		.0592	.1648	.3712	.6440	.8636		.0544	.0768	.1960	.4692	.8128
W		.0572	.2176	.4928	.8228	.9704		.0588	.0876	.2696	.6280	.9236
				σ_u						σ_u		
		0	.2	.3	.4	.5		0	.05	.1	.15	.2
Honda-$T^{\dagger(n)}_{\phi_1}$		.0160	.0172	.0208	.0176	.0168		.0476	.0496	.0816	.1568	.2972
LAD-$T^{\dagger(n)}_{\phi_1}$		.0168	.0176	.0196	.0176	.0180		.0472	.0508	.0852	.1536	.3020
vdW	t_1	.0536	.0884	.1616	.2668	.4344	$\mathcal{S}t_3$	.0568	.0836	.1552	.3332	.6044
t_3		.0568	.1388	.2744	.4920	.7300		.0468	.0708	.1284	.2188	.4196
t_1		.0536	.1648	.3380	.5844	.8300		.0508	.0772	.1688	.3688	.6584
W		.0540	.1180	.2148	.3772	.6104		.0540	.0840	.1660	.3652	.6620

Table 2: Rejection frequencies (out of $M = 2,500$ replications), for $\sigma_u = 0$ (null hypothesis) and various non-zero values of σ_u (alternative hypotheses), with error density g_1 that is Gaussian (ϕ_1), t with 3 degrees of freedom (t_3), Cauchy (t_1), logistic (ℓ_1), skew normal ($\mathcal{SN}$), and skew-t_3 ($\mathcal{S}t_3$), of the OLS-based pseudo-Gaussian test—the Honda test—(Honda-$T^{\dagger(n)}_{\phi_1}$), its LAD-based counterpart (LAD-$T^{\dagger(n)}_{\phi_1}$), and the van der Waerden (vdW), t_ν-score with $\nu = 1, 3$ (t_1, t_3), and Wilcoxon (W) rank tests. The sample size is 500 ($n = 100$ and $T = 5$); see Section 5 for details.

Acknowledgement. We thank the Editor and four referees for their careful and patient reading of the manuscript, and for many suggestions that helped improving the final presentation.

6 Appendix.

6.1 Proof of Proposition 2.1 (ULAN).

The proof consists in cheking that Swensen's (1985) sufficient conditions (1.2)-(1.7) hold. Conditions (1.3)-(1.7) (to be checked, for ULAN, under sequences $\boldsymbol{\vartheta}^{(n)}$) follow more or less routinely from the assumptions made, and the only delicate one actually is condition (1.2). This condition is a direct consequence (see Swensen's Lemma 2) of the quadratic mean differentiability, at any $(\mu, \boldsymbol{\beta}, \sigma^2, 0)$, of

$$(\mu, \boldsymbol{\beta}, \sigma^2, \sigma_u^2) \mapsto \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,\sigma_u^2;f_1}(\mathbf{y}) := \left\{ \frac{1}{\sigma^T} \int_{\mathbb{R}} \prod_{t=1}^{T} f_1 \left(\frac{1}{\sigma} \left(y_t - \mu - \boldsymbol{\beta}'\mathbf{x}_t - \sigma_u u \right) \right) h(u)\, du \right\}^{1/2}$$

with $\mathbf{y} := (y_1, \ldots, y_T)' \in \mathbb{R}^T$ and $\mathbf{x}_t := (x_t^1, x_t^2, \ldots, x_t^K)' \in \mathbb{R}^K$, $t = 1, \ldots, T$. The technical difficulty lies in the fact that $\underline{f}_{\mu,\boldsymbol{\beta},\sigma^2,\sigma_u^2;f_1}$ has the form of a mixture density. Quadratic mean differentiability is established in the following lemma.

Lemma 6.1 *Let Assumptions (B) and (C) hold an fix $f_1 \in \mathcal{F}_A$. Define, for $\mathbf{y} \in \mathbb{R}^T$,*

$$D_\mu \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,0;f_1}(\mathbf{y}) := \frac{1}{2\sigma} \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,0;f_1}(\mathbf{y}) \sum_{t=1}^{T} \phi_{f_1} \left(\frac{y_t - \mu - \boldsymbol{\beta}'\mathbf{x}_t}{\sigma} \right),$$

$$\boldsymbol{D}_\beta \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,0;f_1}(\mathbf{y}) := \frac{1}{2\sigma} \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,0;f_1}(\mathbf{y}) \sum_{t=1}^{T} \phi_{f_1} \left(\frac{y_t - \mu - \boldsymbol{\beta}'\mathbf{x}_t}{\sigma} \right) (\mathbf{K}^{(n)})'\mathbf{x}_t,$$

$$D_{\sigma^2} \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,0;f_1}(\mathbf{y}) := \frac{1}{4\sigma^2} \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,0;f_1}(\mathbf{y}) \sum_{t=1}^{T} \left(\left(\frac{y_t - \mu - \boldsymbol{\beta}'\mathbf{x}_t}{\sigma} \right) \phi_{f_1} \left(\frac{y_t - \mu - \boldsymbol{\beta}'\mathbf{x}_t}{\sigma} \right) - 1 \right),$$

and

$$D_{\sigma_u^2} \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,\sigma_u^2;f_1}(\mathbf{y})_{|\sigma_u^2=0} := \frac{1}{4\sigma^2} \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,0;f_1}(\mathbf{y}) \left[\sum_{t=1}^{T} \psi_{f_1} \left(\frac{y_t - \mu - \boldsymbol{\beta}'\mathbf{x}_t}{\sigma} \right) + \sum_{t=1}^{T} \sum_{1=l\neq t}^{T} \phi_{f_1} \left(\frac{y_t - \mu - \boldsymbol{\beta}'\mathbf{x}_t}{\sigma} \right) \phi_{f_1} \left(\frac{y_t - \mu - \boldsymbol{\beta}'\mathbf{x}_l}{\sigma} \right) \right].$$

Then, as t, s, v, and $r \to 0$,

$$\textit{(i)} \int \left\{ \underline{f}^{1/2}_{\mu+t,\boldsymbol{\beta}+\mathbf{s},\sigma^2+v,r^2;f_1}(\mathbf{y}) - \underline{f}^{1/2}_{\mu+t,\boldsymbol{\beta}+\mathbf{s},\sigma^2+v,0;f_1}(\mathbf{y}) - r^2 D_{\sigma_u^2} \underline{f}^{1/2}_{\mu+t,\boldsymbol{\beta}+\mathbf{s},\sigma^2+v,\sigma_u^2;f_1}(\mathbf{y})_{|\sigma_u^2=0} \right\}^2 d\mathbf{y} = o(r^4),$$

$$\textit{(ii)} \int \left\{ \underline{f}^{1/2}_{\mu+t,\boldsymbol{\beta}+\mathbf{s},\sigma^2+v,0;f_1}(\mathbf{y}) - \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,0;f_1}(\mathbf{y}) - (t, \mathbf{s}', v) \begin{pmatrix} D_\mu \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,0;f_1}(\mathbf{y}) \\ \boldsymbol{D}_{\boldsymbol{\beta}} \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,0;f_1}(\mathbf{y}) \\ D_{\sigma^2} \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,0;f_1}(\mathbf{y}) \end{pmatrix} \right\}^2 d\mathbf{y} = o\left(\|(t, \mathbf{s}', v)\|^2 \right),$$

$$\textit{(iii)} \int \left\{ D_{\sigma_u^2} \underline{f}^{1/2}_{\mu+t,\boldsymbol{\beta}+\mathbf{s},\sigma^2+v,\sigma_u^2;f_1}(\mathbf{y})_{|\sigma_u^2=0} - D_{\sigma_u^2} \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,\sigma_u^2;f_1}(\mathbf{y})_{|\sigma_u^2=0} \right\}^2 d\mathbf{y} = o(1), \textit{ and}$$

$$\textit{(iv)} \int \left\{ \underline{f}^{1/2}_{\mu+t,\boldsymbol{\beta}+\mathbf{s},\sigma^2+v,r^2;f_1}(\mathbf{y}) - \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,0;f_1}(\mathbf{y}) - (t, \mathbf{s}', v, r^2) \begin{pmatrix} D_\mu \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,0;f_1}(\mathbf{y}) \\ \boldsymbol{D}_\beta \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,0;f_1}(\mathbf{y}) \\ D_{\sigma^2} \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,0;f_1}(\mathbf{y}) \\ D_{\sigma_u^2} \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,\sigma_u^2;f_1}(\mathbf{y})_{|\sigma_u^2=0} \end{pmatrix} \right\}^2 d\mathbf{y} = o(\|(t, \mathbf{s}', v, r^2)'\|^2).$$

Proof. (i) Letting $z_t := (y_t - (\mu + t) - (\boldsymbol{\beta} + \mathbf{s})'\mathbf{x}_t) / (\sigma^2 + v)^{1/2}$ and $\mathbf{z} := (z_1, z_2, \ldots, z_T)'$, the left-hand side in (i) takes the form

$$\int_{\mathbb{R}^T} \Bigg[\left(\int_{\mathbb{R}} \prod_{t=1}^{T} f_1 \left(z_t - \frac{r}{(\sigma^2 + v)^{1/2}} u \right) h(u)\, du \right)^{1/2} - \left(\prod_{t=1}^{T} f_1 (z_t) \right)^{1/2}$$
$$- \frac{r^2}{4(\sigma^2 + v)} \left(\prod_{t=1}^{T} f_1 (z_t) \right)^{1/2} \left(\sum_{t=1}^{T} \psi_{f_1} (z_t) + \sum_{t=1}^{T} \sum_{1=l \neq t}^{T} \phi_{f_1} (z_t)\, \phi_{f_1} (z_l) \right) \Bigg]^2 d\mathbf{z}.$$

In order to prove (i), it is thus sufficient to establish differentiability in quadratic mean with respect to $(r/(\sigma^2+v)^{1/2})^2$. This quadratic mean differentiability property, however, is somewhat nonstandard, as it involves the second-order derivatives of the product $\left[\prod_{t=1}^{T} f_1 (z_t)\right]$. As in Akharif and Hallin (2003), the proof is decomposed into three parts.

(a) With the above notation, $y^2 \mapsto L_{\mathbf{z}}(y) := \int_{\mathbb{R}} K_{\mathbf{z}}(u, y) h(u)\, du$, with $K_{\mathbf{z}}(u, y) := \prod_{t=1}^{T} f_1(z_t - yu)$, is absolutely continuous at $y = 0$, with a.e. derivative

$$\begin{aligned} \Psi_{\mathbf{z}}(y) \;&:=\; \frac{1}{2y} \int_{w=0}^{y} \int_{\mathbb{R}} \Big\{ \sum_{t=1}^{T} \ddot{f}_1(z_t - wu) \Big[\prod_{1 \leq l \neq t}^{T} f_1(z_t - wu) \Big] \\ &\quad + \sum_{t=1}^{T} \sum_{1 \leq l \neq t}^{T} \dot{f}_1(z_t - wu) \dot{f}_1(z_l - wu) \Big[\prod_{\substack{m=1 \\ m \neq t, m \neq l}}^{T} f_1(z_m - wu) \Big] \Big\} u^2 h(u)\, du\, dw. \end{aligned} \tag{6.1}$$

We obtain

$$\begin{aligned} L(y) - L(0) \;&=\; \int_{u=-\infty}^{\infty} [K(u, y) - K(u, 0)] h(u)\, du \\ &= \int_{u=-\infty}^{\infty} \int_{a=0}^{y} \dot{K}(u, a)\, da\, h(u)\, du \\ &= \int_{u=-\infty}^{\infty} \int_{a=0}^{y} [\dot{K}(u, a) - \dot{K}(u, 0)] da\ h(u)\ du + \int_{u=-\infty}^{\infty} \int_{a=0}^{y} \dot{K}(u, 0) da\, h(u)\, du \\ &= \int_{u=-\infty}^{\infty} \int_{a=0}^{y} \int_{w=0}^{a} \ddot{K}(u, w) dw\, da\, h(u)\, du, \end{aligned} \tag{6.2}$$

where

$$\begin{aligned} \ddot{K}(u, w) \;&:=\; \sum_{t=1}^{T} \ddot{f}_1(z_t - wu) \Big[\prod_{1 \leq l \neq t}^{T} f_1(z_l - wu) \Big] u^2 \\ &\quad + \sum_{t=1}^{T} \sum_{1 \leq l \neq t}^{T} \dot{f}_1(z_t - wu) \dot{f}_1(z_l - wu) \Big[\prod_{\substack{m=1 \\ m \neq t, m \neq l}}^{T} f_1(z_m - wu) \Big] u^2. \end{aligned}$$

The value (6.1) of the a.e. derivative for $y > 0$ follows. At $y = 0$, the right derivative is defined as the limit, as $y \downarrow 0$, of $[L(y) - L(0)]/y^2$, for which (6.2) yields 0/0. Applying L'Hospital's rule,

$$\lim_{y \downarrow 0} [L(y) - L(0)]/y^2 = \frac{1}{2} \Big\{ \sum_{t=1}^{T} \psi_{f_1}(z_t) + \sum_{t=l}^{T} \sum_{1 \leq l \neq t}^{T} \phi_{f_1}(z_t) \phi_{f_1}(z_l) \Big\} \Big[\prod_{m=1}^{T} f_1(z_m) \Big].$$

(b) It follows that $y^2 \longmapsto s_{\mathbf{z}}(y) := [L_{\mathbf{z}}(y)]^{\frac{1}{2}}$ is also absolutely continuous in a neighborhood of $y = 0$, with a.e. derivative

$$\dot{s}_{\mathbf{z}}(y) = \frac{1}{4y}\int_{w=0}^{y}\int_{\mathbb{R}}\ddot{K}(u,w)h(u)\,du/\Big[\int_{\mathbb{R}}K(u,y)h(u)\,du\Big]^{\frac{1}{2}}dw. \tag{6.3}$$

L'Hospital's rule at $y = 0$ yields

$$\dot{s}_{\mathbf{z}}(0) = \frac{1}{4}\Big(\prod_{l=1}^{T}f_1(z_l)\Big)^{-1/2}\Big\{\sum_{t=1}^{T}\ddot{f}_1(z_t)\Big[\prod_{1\le l\neq t}f_1(z_l)\Big] + \sum_{t=1}^{T}\sum_{1\le l\neq t}^{T}\dot{f}_1(z_t)\dot{f}_1(z_l)\Big[\prod_{\substack{m=1\\ m\neq t, m\neq l}}^{T}f_1(z_m)\Big]\Big\}.$$

Hence, for all $\mathbf{z}$,

$$\lim_{y\to 0} y^{-2}[s_{\mathbf{z}}(y) - s_{\mathbf{z}}(0)] = \dot{s}_{\mathbf{z}}(0). \tag{6.4}$$

(c) The partial quadratic mean differentiablity property to be proved takes the form

$$\lim_{y\to 0}\int_{\mathbb{R}^T}\left\{\frac{1}{y^2}[s_{\mathbf{z}}(y) - s_{\mathbf{z}}(0)] - \frac{1}{4}\dot{s}_{\mathbf{z}}(0)\right\}^2 d\mathbf{z} = 0. \tag{6.5}$$

From (b) above,

$$\left\{\frac{1}{y^2}[s_{\mathbf{z}}(y) - s_{\mathbf{z}}(0)]\right\}^2 = \left(\frac{1}{y^2}\right)^2\left(\int_{\lambda=0}^{y^2}\dot{s}_{\mathbf{z}}(\sqrt{\lambda})d\lambda\right)^2 \le \frac{1}{y^2}\int_{\lambda=0}^{y^2}\left[\dot{s}_{\mathbf{z}}(\sqrt{\lambda})\right]^2 d\lambda,$$

for all $\mathbf{z}$. Fubini's theorem and (6.3) yields

$$\begin{aligned}\int_{\mathbb{R}^T}\left\{\frac{1}{y^2}[s_{\mathbf{z}}(y) - s_{\mathbf{z}}(0)]\right\}^2 d\mathbf{z} &\le \frac{1}{y^2}\int_{\lambda=0}^{y^2}\int_{\mathbb{R}^T}\left[\dot{s}_{\mathbf{z}}(\sqrt{\lambda})\right]^2 d\mathbf{z}d\lambda \\ &= \frac{1}{16y^2}\int_{\lambda=0}^{y^2}\mathcal{I}_{\psi\phi}(f_1;\sqrt{\lambda})d\lambda,\end{aligned} \tag{6.6}$$

with $\mathcal{I}_{\psi\phi}(f_1; y)$ defined in (2.3). The continuity assumption in (B3) implies that this latter quantity converges, as $y \to 0$, to $\mathcal{I}_{\psi\phi}(f_1; 0)/16 = \int_{\mathbb{R}^T}[\dot{s}_{\mathbf{z}}(0)]^2\,d\mathbf{z}$, which together with (6.6), entails that

$$\limsup_{y\to 0}\int_{\mathbb{R}^T}\left\{\frac{1}{y^2}[s_{\mathbf{z}}(y) - s_{\mathbf{z}}(0)]\right\}^2 d\mathbf{z} \le \int_{\mathbb{R}^T}[\dot{s}_{\mathbf{z}}(0)]^2\,d\mathbf{z}. \tag{6.7}$$

In view of Theorem V.I.3 of Hájek and Šidák (1967) [also in Hájek, Šidák and Sen (1999)], (6.4) and (6.7) jointly imply (6.5). This completes the proof of part (i) of the lemma.

(ii) The problem here reduces to the classical case of linear models considered by Swensen (1985).

(iii) First note that, as t, $\mathbf{s} \to 0$,

$$\int_{\mathbb{R}^T}\left\{D_{\sigma_u^2}\underline{f}^{1/2}_{\mu+t,\boldsymbol{\beta}+\mathbf{s},\sigma^2,\sigma_u^2;f_1}(\mathbf{y})_{|\sigma_u^2=0} - D_{\sigma_u^2}\underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,\sigma_u^2;f_1}(\mathbf{y})_{|\sigma_u^2=0}\right\}^2 d\mathbf{y} = o(1).$$

For the perturbation of σ^2, letting $z_t := y_t - \mu - \boldsymbol{\beta}'\mathbf{x}_t$ for $t = 1, \ldots, T$, we have

$$
\begin{aligned}
T \;:=\; & \int_{\mathbb{R}^T} \Big\{ D_{\sigma_u^2} \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2+h,\sigma_u^2;f_1}(\mathbf{y})_{|\sigma_u^2=0} - D_{\sigma_u^2} \underline{f}^{1/2}_{\mu,\boldsymbol{\beta},\sigma^2,\sigma_u^2;f_1}(\mathbf{y})_{|\sigma_u^2=0} \Big\}^2 d\mathbf{y} \\
= \; & \int_{\mathbb{R}^T} \Big\{ \frac{1}{4} \frac{1}{(\sigma^2+h)^{T+4/4}} \Big[\prod_{t=1}^T f_1^{1/2}\Big(\frac{y_t-\mu-\boldsymbol{\beta}'\mathbf{x}_t}{(\sigma^2+h)^{1/2}} \Big) \Big] \\
& \times \Big[\sum_{t=1}^T \psi_{f_1}\Big(\frac{y_t-\mu-\boldsymbol{\beta}'\mathbf{x}_t}{(\sigma^2+h)^{1/2}} \Big) + \sum_{t=1}^T \sum_{1\le l\ne t}^T \phi_{f_1}\Big(\frac{y_t-\mu-\boldsymbol{\beta}'\mathbf{x}_t}{(\sigma^2+h)^{1/2}} \Big) \phi_{f_1}\Big(\frac{y_t-\mu-\boldsymbol{\beta}'\mathbf{x}_t}{(\sigma^2+h)^{1/2}} \Big) \Big] \\
& - \frac{1}{4} \frac{1}{(\sigma^2)^{T+4/4}} \Big[\prod_{t=1}^T f_1^{1/2}\Big(\frac{y_t-\mu-\boldsymbol{\beta}'\mathbf{x}_t}{\sigma} \Big) \Big] \\
& \times \Big[\sum_{t=1}^T \psi_{f_1}\Big(\frac{y_t-\mu-\boldsymbol{\beta}'\mathbf{x}_t}{\sigma} \Big) + \sum_{t=1}^T \sum_{1\le l\ne t}^T \phi_{f_1}\Big(\frac{y_t-\mu-\boldsymbol{\beta}'\mathbf{x}_t}{\sigma} \Big) \phi_{f_1}\Big(\frac{y_t-\mu-x_l\beta}{\sigma} \Big) \Big] \Big\}^2 d\mathbf{y} \\
= \; & \int_{\mathbb{R}^T} \Big\{ \frac{1}{4} \frac{1}{(\sigma^2+h)^{T+4/4}} \Big[\prod_{t=1}^T f_1^{1/2}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) \Big] \\
& \times \Big[\sum_{t=1}^T \psi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) + \sum_{t=1}^T \sum_{1\le l\ne t}^T \phi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) \phi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) \Big] \\
& - \frac{1}{4} \frac{1}{(\sigma^2)^{T+4/4}} \Big[\prod_{t=1}^T f_1^{1/2}\Big(\frac{z_t}{\sigma} \Big) \Big] \Big[\sum_{t=1}^T \psi_{f_1}\Big(\frac{z_t}{\sigma} \Big) + \sum_{t=1}^T \sum_{1\le l\ne t}^T \phi_{f_1}\Big(\frac{z_t}{\sigma} \Big) \phi_{f_1}\Big(\frac{z_l}{\sigma} \Big) \Big] \Big\}^2 d\mathbf{z} \\
= \; & \int_{\mathbb{R}^T} \Big\{ \frac{1}{4} \Big[\frac{1}{(\sigma^2+h)} \prod_{t=1}^T \frac{1}{(\sigma^2+h)^{1/4}} f_1^{1/2}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) - \frac{1}{\sigma^2} \prod_{t=1}^T \frac{1}{\sigma^{1/2}} f_1^{1/2}\Big(\frac{z_t}{\sigma} \Big) \Big] \\
& \times \Big[\sum_{t=1}^T \psi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) + \sum_{t=1}^T \sum_{1\le l\ne t}^T \phi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) \phi_{f_1}\Big(\frac{z_l}{(\sigma^2+h)^{1/2}} \Big) \Big] \\
& + \frac{1}{4} \frac{1}{\sigma^2} \Big[\prod_{t=1}^T \frac{1}{(\sigma^2)^{1/4}} f_1^{1/2}\Big(\frac{z_t}{\sigma} \Big) \Big] \Big\{ \sum_{t=1}^T \Big[\psi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) - \psi_{f_1}\Big(\frac{z_t}{\sigma} \Big) \Big] \\
& + \sum_{t=1}^T \sum_{1\le l\ne t}^T \Big[\phi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) \phi_{f_1}\Big(\frac{z_l}{(\sigma^2+h)^{1/2}} \Big) - \phi_{f_1}\Big(\frac{z_t}{\sigma} \Big) \phi_{f_1}\Big(\frac{z_l}{\sigma} \Big) \Big] \Big\} \Big\}^2 d\mathbf{z} \\
= \; & \int_{\mathbb{R}^T} \Big\{ \frac{1}{4} \Big[\frac{1}{(\sigma^2+h)} - \frac{1}{\sigma^2} \Big] \prod_{t=1}^T \frac{1}{(\sigma^2+h)^{1/4}} f_1^{1/2}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) \\
& \qquad \times \Big[\sum_{t=1}^T \psi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) + \sum_{t=1}^T \sum_{1\le l\ne t}^T \phi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) \phi_{f_1}\Big(\frac{z_l}{(\sigma^2+h)^{1/2}} \Big) \Big] \\
& + \frac{1}{4} \Big[\frac{1}{\sigma^2} \prod_{t=1}^T \frac{1}{(\sigma^2+h)^{1/4}} f_1^{1/2}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) - \frac{1}{\sigma^2} \prod_{t=1}^T \frac{1}{\sigma^{1/2}} f_1^{1/2}\Big(\frac{z_t}{\sigma} \Big) \Big] \\
& \qquad \times \Big[\sum_{t=1}^T \psi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) + \sum_{t=1}^T \sum_{1\le l\ne t}^T \phi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) \phi_{f_1}\Big(\frac{z_l}{(\sigma^2+h)^{1/2}} \Big) \Big] \\
& + \frac{1}{4\sigma^2} \Big[\prod_{t=1}^T \frac{1}{(\sigma^2)^{1/4}} f_1^{1/2}\Big(\frac{z_t}{\sigma} \Big) \Big] \Big\{ \sum_{t=1}^T \Big[\psi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) - \psi_{f_1}\Big(\frac{z_t}{\sigma} \Big) \Big]
\end{aligned}
$$

$$
\begin{aligned}
& + \sum_{t=1}^{T} \sum_{1 \leq l \neq t}^{T} \Big[\phi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) \phi_{f_1}\Big(\frac{z_l}{(\sigma^2+h)^{1/2}} \Big) - \phi_{f_1}\Big(\frac{z_t}{\sigma} \Big) \phi_{f_1}\Big(\frac{z_l}{\sigma} \Big) \Big] \Big\} \Big\}^2 d\mathbf{z} \\
\leq \;\; & C\,(T_1 + T_2 + T_3),
\end{aligned}
$$

where

$$
\begin{aligned}
T_1 \;\; := \;\; & \int_{\mathbb{R}^T} \Big\{ \frac{1}{4} \Big[\frac{1}{(\sigma^2+h)} - \frac{1}{\sigma^2} \Big] \Big[\prod_{t=1}^{T} \frac{1}{(\sigma^2+h)^{1/4}} f_1^{1/2}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) \Big] \\
& \times \Big[\sum_{t=1}^{T} \psi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) + \sum_{t=1}^{T} \sum_{1 \leq l \neq t}^{T} \phi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) \phi_{f_1}\Big(\frac{z_l}{(\sigma^2+h)^{1/2}} \Big) \Big] \Big\}^2 d\mathbf{z}, \\
T_2 \;\; := \;\; & \int_{\mathbb{R}^T} \Big\{ \frac{1}{4\sigma^2} \Big[\prod_{t=1}^{T} \frac{1}{(\sigma^2+h)^{1/4}} f_1^{1/2}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) - \prod_{t=1}^{T} \frac{1}{\sigma^{1/2}} f_1^{1/2}\Big(\frac{z_t}{\sigma} \Big) \Big] \\
& \times \Big[\sum_{t=1}^{T} \psi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) + \sum_{t=1}^{T} \sum_{1 \leq l \neq t}^{T} \phi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) \phi_{f_1}\Big(\frac{z_l}{(\sigma^2+h)^{1/2}} \Big) \Big] \Big\}^2 d\mathbf{z},
\end{aligned}
$$

and

$$
\begin{aligned}
T_3 \;\; := \;\; & \int_{\mathbb{R}^T} \Big\{ \frac{1}{4\sigma^2} \Big[\prod_{t=1}^{T} \frac{1}{(\sigma^2)^{1/4}} f_1^{1/2}\Big(\frac{z_t}{\sigma} \Big) \Big] \Big\{ \sum_{t=1}^{T} \Big[\psi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) - \psi_{f_1}\Big(\frac{z_t}{\sigma} \Big) \Big] \\
& + \sum_{t=1}^{T} \sum_{1 \leq l \neq t}^{T} \Big[\phi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) \phi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) - \phi_{f_1}\Big(\frac{z_t}{\sigma} \Big) \phi_{f_1}\Big(\frac{z_t}{\sigma} \Big) \Big] \Big\} \Big\}^2 d\mathbf{z}.
\end{aligned}
$$

Clearly, $T_1 = O\big(\big[(\sigma^2+h)^{-1} - \sigma^{-2}\big]^2 (\mathcal{I}_\psi(f_1) + \mathcal{I}^2_\phi(f_1))\big)$, which implies that $T_1 = o(1)$, as $h \to 0$. Turning to T_2, we have

$$
\begin{aligned}
T_2 \;\; = \;\; & \int_{\mathbb{R}^T} \Big\{ \frac{1}{4\sigma^2} \Big\{ \sum_{t=1}^{T} \Big[\frac{1}{(\sigma^2+h)^{1/4}} f_1^{1/2}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) - \frac{1}{\sigma^{1/2}} f_1^{1/2}\Big(\frac{z_t}{\sigma} \Big) \Big] \\
& \times \prod_{k=1}^{t-1} \frac{1}{(\sigma^2+h)^{1/4}} f_1^{1/2}\Big(\frac{z_k}{(\sigma^2+h)^{1/2}} \Big) \prod_{k=t-1}^{T} \frac{1}{(\sigma^2)^{1/4}} f_1^{1/2}\Big(\frac{z_k}{\sigma} \Big) \Big\} \\
& \times \Big\{ \sum_{t=1}^{T} \psi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) + \sum_{t=1}^{T} \sum_{1 \leq l \neq t}^{T} \phi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) \phi_{f_1}\Big(\frac{z_l}{(\sigma^2+h)^{1/2}} \Big) \Big\} \Big\}^2 d\mathbf{z} \\
= \;\; & \int_{\mathbb{R}^T} \Big\{ \frac{1}{4\sigma^2} \Big\{ \sum_{t=1}^{T} \sum_{l=1}^{T} \Big[\frac{1}{(\sigma^2+h)^{1/4}} f_1^{1/2}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) - \frac{1}{\sigma^{1/2}} f_1^{1/2}\Big(\frac{z_t}{\sigma} \Big) \Big] \\
& \times \psi_{f_1}\Big(\frac{z_l}{(\sigma^2+h)^{1/2}} \Big) \prod_{k=1}^{t-1} \frac{1}{(\sigma^2+h)^{1/4}} f_1^{1/2}\Big(\frac{z_k}{(\sigma^2+h)^{1/2}} \Big) \prod_{k=t-1}^{T} \frac{1}{(\sigma^2)^{1/4}} f_1^{1/2}\Big(\frac{z_k}{\sigma} \Big) \Big\} \\
& + \frac{1}{4\sigma^2} \sum_{t=1}^{T} \sum_{l=1}^{T} \sum_{\substack{m=1 \\ m \neq l}}^{T} \Big[\frac{1}{(\sigma^2+h)^{1/4}} f_1^{1/2}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}} \Big) - \frac{1}{\sigma^{1/2}} f_1^{1/2}\Big(\frac{z_t}{\sigma} \Big) \Big] \phi_{f_1}\Big(\frac{z_l}{(\sigma^2+h)^{1/2}} \Big) \\
& \times \phi_{f_1}\Big(\frac{z_m}{(\sigma^2+h)^{1/2}} \Big) \prod_{k=1}^{t-1} \frac{1}{(\sigma^2+h)^{1/4}} f_1^{1/2}\Big(\frac{z_k}{(\sigma^2+h)^{1/2}} \Big) \prod_{k=t-1}^{T} \frac{1}{(\sigma^2)^{1/4}} f_1^{1/2}\Big(\frac{z_k}{\sigma} \Big) \Big\}^2 d\mathbf{z} \\
\leq \;\; & C_1 (T_2^1 + T_2^2),
\end{aligned}
$$

where

$$T_2^1 \;:=\; \int_{\mathbb{R}^T} \Big\{\frac{1}{4\sigma^2}\Big\{\sum_{t=1}^{T}\sum_{l=1}^{T}\Big[\frac{1}{(\sigma^2+h)^{1/4}} f_1^{1/2}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}}\Big) - \frac{1}{\sigma^{1/2}} f_1^{1/2}\Big(\frac{z_t}{\sigma}\Big)\Big]$$
$$\times \psi_{f_1}\Big(\frac{z_l}{(\sigma^2+h)^{1/2}}\Big)\prod_{k=1}^{t-1}\frac{1}{(\sigma^2+h)^{1/4}} f_1^{1/2}\Big(\frac{z_k}{(\sigma^2+h)^{1/2}}\Big)\prod_{k=t-1}^{T}\frac{1}{(\sigma^2)^{1/4}} f_1^{1/2}\Big(\frac{z_k}{\sigma}\Big)\Big\}\Big\}^2 d\mathbf{z}$$

and

$$T_2^2 \;:=\; \int_{\mathbb{R}^T} \Big\{\frac{1}{4\sigma^2}\sum_{t=1}^{T}\sum_{l=1}^{T}\sum_{\substack{m=1\\ m\neq l}}^{T}\Big[\frac{1}{(\sigma^2+h)^{1/4}} f_1^{1/2}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}}\Big) - \frac{1}{\sigma^{1/2}} f_1^{1/2}\Big(\frac{z_t}{\sigma}\Big)\Big]\phi_{f_1}\Big(\frac{z_l}{(\sigma^2+h)^{1/2}}\Big)$$
$$\times \phi_{f_1}\Big(\frac{z_m}{(\sigma^2+h)^{1/2}}\Big)\prod_{k=1}^{t-1}\frac{1}{(\sigma^2+h)^{1/4}} f_1^{1/2}\Big(\frac{z_k}{(\sigma^2+h)^{1/2}}\Big)\prod_{k=t-1}^{T}\frac{1}{(\sigma^2)^{1/4}} f_1^{1/2}\Big(\frac{z_k}{\sigma}\Big)\Big\}^2 d\mathbf{z}.$$

Then, in order to prove that $T_2 = o(1)$, it is clearly sufficient to show that T_2^1 and T_2^2 are o(1) as $h \to 0$. We start with T_2^1, which is bounded by $B\,(T_2^{11}+T_2^{12}+T_2^{13})$ for some constant $B > 0$.

$$T_2^{11} := \int_{\mathbb{R}^T}\Big\{\sum_{t=1}^{T}\Big[\frac{1}{(\sigma^2+h)^{1/4}} f_1^{1/2}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}}\Big)\psi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}}\Big) - \frac{1}{\sigma^{1/2}} f_1^{1/2}\left(\frac{z_t}{\sigma}\right)\psi_{f_1}\left(\frac{z_t}{\sigma}\right)\Big]$$
$$\times\prod_{k=1}^{t-1}\frac{1}{(\sigma^2+h)^{1/4}} f_1^{1/2}\Big(\frac{z_k}{(\sigma^2+h)^{1/2}}\Big)\prod_{k=t-1}^{T}\frac{1}{(\sigma^2)^{1/4}} f_1^{1/2}\left(\frac{z_k}{\sigma}\right)\Big\}^2 d\mathbf{z}$$
$$\le B_1\int_{\mathbb{R}}\Big\{e^{\frac{1}{2}\left[u-\ln(1+\frac{h}{\sigma^2})^{1/2}\right]} f_1^{1/2}\Big(e^{u-\ln(1+\frac{h}{\sigma^2})^{1/2}}\Big)\psi_{f_1}\Big(e^{u-\ln(1+\frac{h}{\sigma^2})^{1/2}}\Big)$$
$$-e^{\frac{1}{2}u} f_1^{1/2}(e^u)\,\psi_{f_1}(e^u)\Big\}^2 du, \qquad (6.8)$$

$$T_2^{12} := \int_{\mathbb{R}}\Big\{\sum_{t=1}^{T}\Big[\psi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}}\Big) - \psi_{f_1}\Big(\frac{z_t}{\sigma}\Big)\Big]\frac{1}{\sigma^{1/2}} f_1^{1/2}\Big(\frac{z_t}{\sigma}\Big)$$
$$\times\prod_{k=1}^{t-1}\frac{1}{(\sigma^2+h)^{1/4}} f_1^{1/2}\Big(\frac{z_k}{(\sigma^2+h)^{1/2}}\Big)\prod_{k=t-1}^{T}\frac{1}{(\sigma^2)^{1/4}} f_1^{1/2}\Big(\frac{z_k}{\sigma}\Big)\Big\}^2 d\mathbf{z}$$
$$\le B_2\int_{\mathbb{R}}\Big\{\psi_{f_1}\Big(e^{u-\ln(1+\frac{h}{\sigma^2})^{1/2}}\Big) - \psi_{f_1}(e^u)\Big\}^2 e^u f_1(e^u)du \qquad (6.9)$$

and

$$T_2^{13} \;:=\; \int_{\mathbb{R}^T}\Big\{2\sum_{t=2}^{T}\sum_{l=1}^{t-1}\Big[\frac{1}{(\sigma^2+h)^{1/4}} f_1^{1/2}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}}\Big) - \frac{1}{\sigma^{1/2}} f_1^{1/2}\Big(\frac{z_t}{\sigma}\Big)\Big]$$
$$\times\psi_{f_1}\Big(\frac{z_l}{(\sigma^2+h)^{1/2}}\Big)\prod_{k=1}^{t-1}\frac{1}{(\sigma^2+h)^{1/4}} f_1^{1/2}\Big(\frac{z_k}{(\sigma^2+h)^{1/2}}\Big)\prod_{k=t-1}^{T}\frac{1}{(\sigma^2)^{1/4}} f_1^{1/2}\Big(\frac{z_k}{\sigma}\Big)\Big\}^2 d\mathbf{z}$$
$$\le\; B_3\int_{\mathbb{R}}\Big\{e^{\frac{1}{2}\left[u-\ln(1+\frac{h}{\sigma^2})^{1/2}\right]} f_1^{1/2}\Big(e^{u-\ln(1+\frac{h}{\sigma^2})^{1/2}}\Big) - e^{\frac{1}{2}u} f_1^{1/2}(e^u)\Big\}^2 du\times\mathcal{I}_\psi(f_1). \qquad (6.10)$$

Since $e^{\frac{1}{2}u} f_1^{1/2}(e^u)$, $e^{\frac{1}{2}u} f_1^{1/2}(e^u)\psi_{f_1}(e^u)$ and $\psi_{f_1}(e^u)$ are square integrable, quadratic mean continuity implies that the integrals in (6.8), (6.9), and (6.10) are $o(1)$ as $h \to 0$.

Similarly, it easily shown that

$$T_2^2 \le C\Big[\int_{\mathbb{R}}\Big\{e^{\frac{1}{2}[u-\ln(1+\frac{h}{\sigma^2})^{1/2}]} f_1^{1/2}\Big(e^{u-\ln(1+\frac{h}{\sigma^2})^{1/2}}\Big)\phi_{f_1}\Big(e^{u-\ln(1+\frac{h}{\sigma^2})^{1/2}}\Big) - e^{\frac{1}{2}u} f_1^{1/2}(e^u)\psi_{f_1}(e^u)\Big\}^2 du$$

$$\times \mathcal{I}_\phi(f_1) + \int_{\mathbb{R}}\Big\{\phi_{f_1}\Big(e^{u-\ln(1+\frac{h}{\sigma^2})^{1/2}}\Big) - \phi_{f_1}(e^u)\Big\}^2 e^u f_1(e^u)du \times \mathcal{I}_\phi(f_1)$$

$$+\int_{\mathbb{R}}\Big\{e^{\frac{1}{2}[u-\ln(1+\frac{h}{\sigma^2})^{1/2}]} f_1^{1/2}\Big(e^{u-\ln(1+\frac{h}{\sigma^2})^{1/2}}\Big) - e^{\frac{1}{2}u} f_1^{1/2}(e^u)\Big\}^2 du \times \mathcal{I}_\phi^2(f_1)\Big] = o(1), \text{ as } h \to 0,$$

since $e^{\frac{1}{2}u} f_1^{1/2}(e^u)$, $e^{\frac{1}{2}u} f_1^{1/2}(e^u)\psi_{f_1}(e^u)$ and $\phi_{f_1}(e^u)$ are square integrable.

As for T_3, note that $T_3 \le D\,(T_3^1 + T_3^2)$ where

$$T_3^1 := \int_{\mathbb{R}^T}\Big\{\frac{1}{4\sigma^2}\Big[\prod_{t=1}^T \frac{1}{(\sigma^2)^{1/4}} f_1^{1/2}\Big(\frac{z_t}{\sigma}\Big)\Big]\Big\{\sum_{t=1}^T \Big[\psi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}}\Big) - \psi_{f_1}\Big(\frac{z_t}{\sigma}\Big)\Big]\Big\}\Big\}^2 d\mathbf{z}$$

$$\le D_1 \int_{\mathbb{R}}\Big(\psi_{f_1}\Big\{e^{u-\ln(1+\frac{h}{\sigma^2})^{1/2}}\Big) - \psi_{f_1}(e^u)\Big\}^2 e^u f_1(e^u)du = o(1), \text{ as } h \to 0,$$

and

$$T_3^2 := \int_{\mathbb{R}^T}\Big\{\frac{1}{4\sigma^2}\Big[\prod_{t=1}^T \frac{1}{(\sigma^2)^{1/4}} f_1^{1/2}\Big(\frac{z_t}{\sigma}\Big)\Big]\Big\{\sum_{t=1}^T \sum_{1=l\ne t}^T \Big[\phi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}}\Big)\phi_{f_1}\Big(\frac{z_t}{(\sigma^2+h)^{1/2}}\Big)$$

$$-\phi_{f_1}\Big(\frac{z_t}{\sigma}\Big)\phi_{f_1}\Big(\frac{z_t}{\sigma}\Big)\Big]\Big\}\Big\}^2 d\mathbf{z}$$

$$\le D_2 \int_{\mathbb{R}}\Big\{\phi_{f_1}\Big(e^{u-\ln(1+\frac{h}{\sigma^2})^{1/2}}\Big) - \phi_{f_1}(e^u)\Big\}^2 e^u f_1(e^u)du = o(1), \text{ as } h \to 0,$$

since $\phi_{f_1}(e^u)$ and $\psi_{f_1}(e^u)$ are square integrable. This completes the proof of Lemma 6.1, hence that of Proposition 2.1. □

6.2 Proof of Proposition 4.1 (Asymptotic representation).

Proof. The proof of Part (i) of the proposition follows along the same lines as in Section 4.1 of Hallin, Ingenbleek, and Puri (1985), and therefore is omitted. Part (ii) is obtained by direct computation. As for Part (iii), under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g_1}$, the result straightforwardly follows from classical central limit theorems. On the other hand, it is easy to see that, still under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g_1,h_1}$, $\boldsymbol{\Delta}^{*(n)}_{f_1,g_1;4}(\boldsymbol{\vartheta})$ and the log-likelihood $\Lambda^{(n)}_{\boldsymbol{\vartheta}+n^{-1/2}\boldsymbol{\nu}^{(n)}\boldsymbol{\tau}/\boldsymbol{\vartheta};g_1}$ are jointly multinormal; the desired result then follows from a routine application of Le Cam's Third Lemma. □

6.3 Asymptotic linearity.

In this section, we regroup the asymptotic linearity results leading to Propositions 3.2 (for the pseudo-Gaussian statistic; see (3.17) for the definition of $T^{\dagger(n)}_{\mathcal{N};g}$) and 4.1 (for rank-based statistics). In the pseudo-Gaussian case, asymptotic normality is needed for perturbations of $\boldsymbol{\beta}$ only, since the linearity of Gaussian scores allows for a direct control of the perturbations of μ and σ^2 through the continuous mapping theorem.

Lemma 6.2 *Let Assumptions (B) and (C) hold; fix $\boldsymbol{\theta} \in \mathbb{R}^{K+1}$, $\boldsymbol{\beta} \in \mathbb{R}^K$, $\sigma^2 > 0$ and $g_1 \in \mathcal{F}_A^2$.*

(i) For any bounded sequence $\mathbf{b}^{(n)} \in \mathbb{R}^K$,

$$T^{\dagger(n)}_{\mathcal{N};g}(\boldsymbol{\beta} + n^{-1/2}\mathbf{K}^{(n)}\mathbf{b}^{(n)}) - T^{\dagger(n)}_{\mathcal{N};g}(\boldsymbol{\beta}) = o_{\mathrm{P}}(1), \tag{6.11}$$

under $\mathrm{P}^{(n)}_{\mu,\boldsymbol{\beta},\sigma^2,0;g_1}$, as $n \to \infty$ with fixed T.

(ii) Assuming that $\hat{\boldsymbol{\beta}}_{\#}^{(n)}$ *satisfies (D1) and (D2),* $T_{\mathcal{N};g}^{\dagger(n)}(\hat{\boldsymbol{\beta}}_{\#}^{(n)}) - T_{\mathcal{N};g}^{\dagger(n)}(\boldsymbol{\beta}) = o_{\mathrm{P}}(1)$*, under* $\mathrm{P}^{(n)}_{\mu,\boldsymbol{\beta},\sigma^2,0;g_1}$*, as* $n\to\infty$ *with fixed* T*.*

Proof of Lemma 6.2. (i) Note that, for $i = 1,\dots,n$ and $t = 1,\dots,T$,

$$W_{it}(\boldsymbol{\beta} + n^{-1/2}\mathbf{K}^{(n)}\,\mathbf{b}^{(n)}) - W_{it}(\boldsymbol{\beta}) = -n^{-1/2}\mathbf{b}^{(n)\prime}\,\mathbf{K}'(\mathbf{D}^{(n)})^{-1/2}\mathbf{x}_{it} =: -\alpha_{it}^{(n)},$$

where, in view of the fact that $\sum_{i=1}^n\sum_{t=1}^T \mathbf{x}_{it} = \mathbf{0}$, $\sum_{i=1}^n\sum_{t=1}^T \alpha_{it}^{(n)} = 0$. We thus have

$$\begin{aligned}
T_{\mathcal{N};g}^{\dagger(n)}(\boldsymbol{\beta} + n^{-1/2}\mathbf{K}^{(n)}\,\mathbf{b}^{(n)}) - T_{\mathcal{N};g}^{\dagger(n)}(\boldsymbol{\beta}) = {}& \frac{1}{\sigma^2(g)\sqrt{2nT(T-1)}}\\
&\times\sum_{i=1}^n\Big\{\sum_{t=1}^T\sum_{l\neq t=1}^T\Big(W_{it}(\boldsymbol{\beta} + n^{-1/2}\mathbf{K}^{(n)}\,\mathbf{b}^{(n)}) - W_{it}(\boldsymbol{\beta})\Big)W_{il}(\boldsymbol{\beta} + n^{-1/2}\mathbf{K}^{(n)}\,\mathbf{b}^{(n)})\\
&\qquad - \sum_{t=1}^T\sum_{l\neq t=1}^T W_{it}(\boldsymbol{\beta})\Big(W_{il}(\boldsymbol{\beta}) - W_{il}(\boldsymbol{\beta} + n^{-1/2}\mathbf{K}^{(n)}\,\mathbf{b}^{(n)})\Big)\Big\}\\
= {}& \frac{1}{\sigma^2(g)\sqrt{2nT(T-1)}}\sum_{i=1}^n\left\{\sum_{t=1}^T\sum_{l\neq t=1}^T -\alpha_{it}^{(n)}(W_{il}(\boldsymbol{\beta}) - \alpha_{il}^{(n)}) - \sum_{t=1}^T\sum_{l\neq t=1}^T W_{it}(\boldsymbol{\beta})\alpha_{il}^{(n)}\right\}\\
= {}& \frac{1}{\sigma^2(g)\sqrt{2nT(T-1)}}\sum_{i=1}^n\sum_{t=1}^T\sum_{l\neq t=1}^T \alpha_{it}^{(n)}\alpha_{il}^{(n)} - \frac{2}{\sigma^2(g)\sqrt{2nT(T-1)}}\sum_{i=1}^n\sum_{t=1}^T\sum_{l\neq t=1}^T W_{it}(\boldsymbol{\beta})\alpha_{il}^{(n)}\\
=: {}& \frac{1}{\sigma^2(g)\sqrt{2T(T-1)}}\mathrm{U}^{(n)} - \frac{2}{\sigma^2(g)\sqrt{2T(T-1)}}\mathrm{V}^{(n)}.
\end{aligned}$$

In order to prove (i), it is sufficient to show that $\mathrm{U}^{(n)}$ and $\mathrm{V}^{(n)}$ are $o_{\mathrm{P}}(1)$ under $\mathrm{P}^{(n)}_{\mu,\boldsymbol{\beta},\sigma^2,0;g_1}$. Starting with $\mathrm{U}^{(n)}$, we have

$$\begin{aligned}
\left|\sum_{i=1}^n\sum_{t=1}^T\sum_{l\neq t=1}^T \alpha_{it}^{(n)}\alpha_{il}^{(n)}\right| &\leq (T-1)\sum_{i=1}^n\sum_{t=1}^T \alpha_{it}^{2(n)}\\
&= (T-1)\,n^{-1}\sum_{i=1}^n\sum_{t=1}^T \mathbf{b}^{(n)\prime}\,\mathbf{K}'(\mathbf{D}^{(n)})^{-1/2}\mathbf{x}_{it}\mathbf{x}_{it}'(\mathbf{D}^{(n)})^{-1/2}\mathbf{K}\,\mathbf{b}^{(n)}\\
&= (T-1)\,\mathbf{b}^{(n)\prime}\,\mathbf{K}'\mathbf{R}^{(n)}\mathbf{K}\,\mathbf{b}^{(n)} = O(1) \qquad (6.12)
\end{aligned}$$

since $\mathbf{K}'\mathbf{R}^{(n)}\mathbf{K}$ converges to an identity matrix, while $\mathbf{b}^{(n)}$ is bounded. The desired result that $\mathrm{U}^{(n)}$ is $O(n^{-1/2})$, hence $o(1)$, follows.

As for $\mathrm{V}^{(n)}$, it suffices to show that $\mathrm{E}[(\mathrm{V}^{(n)})^2] = o(1)$ under $\mathrm{P}^{(n)}_{\mu,\boldsymbol{\beta},\sigma^2,0;g_1}$. Still because of the fact that $\sum_{i=1}^n\sum_{t=1}^T \alpha_{it}^{(n)} = 0$,

$$\mathrm{V}^{(n)} = \frac{\sigma}{\sqrt{n}}\sum_{i=1}^n\sum_{t=1}^T\sum_{l\neq t=1}^T \alpha_{it}^{(n)}(Z_{il}(\boldsymbol{\theta},\sigma^2) - m_1^{(n)}) = \frac{\sigma}{\sqrt{n}}\sum_{i=1}^n\sum_{t=1}^T\sum_{l\neq t=1}^T \alpha_{it}^{(n)}(Z_{il}(\boldsymbol{\theta},\sigma^2) - \mu_1(g_1)) =: \frac{\sigma}{\sqrt{n}}\sum_{i=1}^n \mathrm{V}_i^{(n)}$$

where, for $i = 1, \ldots, n$, the $\mathrm{V}_i^{(n)}$'s, under $\mathrm{P}^{(n)}_{\mu,\boldsymbol{\beta},\sigma^2,0;g_1}$, are i.i.d. with mean zero. Hence,

$$\begin{aligned}
\mathrm{E}[(\mathrm{V}^{(n)})^2] &= \frac{\sigma^2}{n}\sum_{i=1}^n \mathrm{E}\Big[(\mathrm{V}_i^{(n)})^2\Big] = \frac{\sigma^2}{n}\sum_{i=1}^n \mathrm{E}\Big[\Big(\sum_{t=1}^T \sum_{l\neq t=1}^T \alpha_{it}(Z_{il}(\boldsymbol{\theta},\sigma^2)-\mu_1(g_1))\Big)^2\Big] \\
&= \frac{\sigma^2}{n}\sum_{i=1}^n\sum_{t=1}^T \mathrm{E}\Big[\Big(Z_{it}(\boldsymbol{\theta},\sigma^2)-\mu_1(g_1)\Big)^2\Big]\Big(\sum_{l\neq t=1}^T \alpha_{il}\Big)^2 = \frac{\sigma^2\sigma^2(g_1)}{n}\sum_{i=1}^n\sum_{t=1}^T\Big(\sum_{l\neq t=1}^T \alpha_{il}\Big)^2 \\
&\leq \frac{\sigma^2\sigma^2(g_1)}{n}\sum_{i=1}^n\sum_{t=1}^T (T-1)\sum_{l\neq t=1}^T \alpha_{il}^2 = \frac{\sigma^2\sigma^2(g_1)}{n}(T-1)^2\sum_{i=1}^n\sum_{t=1}^T \alpha_{it}^2,
\end{aligned}$$

a quantity which, in view of (6.12), is $O(n^{-1})$. The desired result that $\mathrm{E}[(\mathrm{V}^{(n)})^2] = o(1)$ (under $\mathrm{P}^{(n)}_{\mu,\boldsymbol{\beta},\sigma^2,0;g_1}$, as $n\to\infty$ with fixed T) follows.

Part (ii) of the lemma then readily follows from part (i) and an application of Lemma 4.4 in Kreiss (1987) (in which asymptotic discreteness plays the major role). □

The corresponding result in the rank-based context is the following; again, we only have to care for the perturbations of $\boldsymbol{\beta}$, since μ and σ^2 have no impact on the ranks, hence on the test statistics $\underset{\sim}{T}{}^{*(n)}_{f_1}(\boldsymbol{\beta})$.

Lemma 6.3 *Let Assumptions (B) and (C) hold; fix* $\boldsymbol{\theta}\in\mathbb{R}^{K+1}$, $\sigma^2>0$, $f_1\in\underset{\sim}{\mathcal{F}}_A$, *and* $g_1\in\mathcal{F}_A$.

(i) For any bounded sequence $\mathbf{b}^{(n)}\in\mathbb{R}^K$,

$$\underset{\sim}{\Delta}{}^{*(n)}_{f_1;4}(\boldsymbol{\beta}+n^{-1/2}\mathbf{K}^{(n)}\,\mathbf{b}^{(n)},\sigma^2) - \underset{\sim}{\Delta}{}^{*(n)}_{f_1;4}(\boldsymbol{\beta},\sigma^2) = o_{\mathrm{P}}(1), \tag{6.13}$$

under $\mathrm{P}^{(n)}_{\mu,\boldsymbol{\beta},\sigma^2,0;g_1}$, *as* $n\to\infty$ *with fixed* T.

(ii) Assuming that $\hat{\boldsymbol{\beta}}^{(n)}_{\#}$ *satisfies (D1) and (D2),* $\underset{\sim}{\Delta}{}^{*(n)}_{f_1;4}(\hat{\boldsymbol{\beta}}^{(n)}_{\#},\sigma^2)-\underset{\sim}{\Delta}{}^{*(n)}_{f_1;4}(\boldsymbol{\beta},\sigma^2)=o_{\mathrm{P}}(1)$, *under* $\mathrm{P}^{(n)}_{\mu,\boldsymbol{\beta},\sigma^2,0;g_1}$, *as* $n\to\infty$ *with fixed* T.

Proof of Lemma 6.3. Part (ii) of the Lemma again follows from part (i) via Lemma 4.4 in Kreiss (1987). We thus focus on the proof of (6.13). Throughout the proof, we let $\boldsymbol{\theta}^{(n)} := \boldsymbol{\theta}+\boldsymbol{\nu}_1^{(n)}\,\mathbf{t}^{(n)}$. Accordingly, let $Z^0_{it} := \sigma^{-1}(Y_{it}-\underset{\sim}{x}'_{it}\,\boldsymbol{\theta})$ and $Z^n_{it} := \sigma^{-1}(Y_{it}-\underset{\sim}{x}'_{it}\,\boldsymbol{\theta}^{(n)})$. Proposition 4.1 implies that $\underset{\sim}{\Delta}{}^{*(n)}_{f_1;4}(\boldsymbol{\beta},\sigma^2)-\Delta^{*(n)}_{f_1,g_1;4}(\boldsymbol{\theta},\sigma^2)$ is $o_{\mathrm{P}}(1)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,0;g_1}$. Similarly, $\underset{\sim}{\Delta}{}^{*(n)}_{f_1;4}(\boldsymbol{\beta}^{(n)},\sigma^2)-\Delta^{*(n)}_{f_1,g_1;4}(\boldsymbol{\theta}^{(n)},\sigma^2)$ is $o_{\mathrm{P}}(1)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}^{(n)},\sigma^2,0;g_1}$ – hence, from contiguity, also under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g_1}$. Consequently, the left-hand side in (6.13) is asymptotically equivalent, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,0;g_1}$, to

$$\begin{aligned}
\mathrm{C}^{(n)}(\boldsymbol{\theta},\sigma^2) &:= \Delta^{*(n)}_{f_1,g_1;4}(\boldsymbol{\theta}^{(n)},\sigma^2,0)-\Delta^{*(n)}_{f_1,g_1;4}(\boldsymbol{\theta},\sigma^2,0) \\
&= n^{-1/2}\frac{1}{2\sigma^2}\sum_{i=1}^n\Big\{\sum_{t=1}^T\sum_{1=l\neq t}^T \phi_{f_1}\left(F_1^{-1}(G_1(Z^n_{it}))\right)\phi_{f_1}\left(F_1^{-1}(G_1(Z^n_{il}))\right) \\
&\qquad -\sum_{t=1}^T\sum_{1=l\neq t}^T \phi_{f_1}\left(F_1^{-1}(G_1(Z^0_{it}))\right)\phi_{f_1}\left(F_1^{-1}(G_1(Z^0_{il}))\right)\Big\},
\end{aligned}$$

and we need only to prove that $\mathrm{C}^{(n)}(\boldsymbol{\theta},\sigma^2)$ is $o_{\mathrm{P}}(1)$.

Let $K_{f_1}(u) := \phi_{f_1}(F_1^{-1}(u))$ and consider the following truncation: for all $\ell \in \mathbb{N}_0$, define

$$K_{f_1}^{(\ell)}(u) \;:=\; K_{f_1}\left(\frac{2}{\ell}\right)\ell(u-\frac{1}{\ell})I_{[\frac{1}{\ell}<u\leq\frac{2}{\ell}]} + K_{f_1}(u)I_{[\frac{2}{\ell}<u\leq 1-\frac{2}{l}]} + K_{f_1}\left(1-\frac{2}{\ell}\right)\ell\left(\left(1-\frac{1}{\ell}\right)-u\right)I_{[1-\frac{2}{\ell}<u\leq 1-\frac{1}{\ell}]}.$$

Continuity of $u \mapsto K_{f_1}(u)$ implies the continuity of $u \mapsto K_{f_1}^{(\ell)}(u)$ on the interval $(0,1)$. It follows that the truncated scores $K_{f_1}^{(\ell)}$ are bounded for all $\ell \in \mathbb{N}_0$. Moreover, since we have assumed that K_{f_1} is monotone increasing, there exists some L such that $|K_{f_1}^{(\ell)}(u)| \leq |K_{f_1}(u)|$ for all $u \in (0,1)$ and all $l \geq L$.

One shows easily that $\mathrm{C}^{(n)}(\boldsymbol{\theta}, \sigma^2)$ decomposes into $\mathrm{D}_1^{(n;\ell)} + \mathrm{D}_2^{(n;\ell)} - \mathrm{R}_1^{(n;\ell)} + \mathrm{R}_2^{(n;\ell)}$, where, denoting by E_0 and Var_0, respectively, expectation and variance under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,0;g_1}$,

$$\begin{aligned}
\mathrm{D}_1^{(n;\ell)} \;:=\; & \frac{n^{-1/2}}{2\sigma^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{1=l\neq t}^{T} K_{f_1}^{(\ell)}(G_1(Z_{it}^n))K_{f_1}^{(\ell)}(G_1(Z_{il}^n)) \\
& -\frac{n^{-1/2}}{2\sigma^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{1=l\neq t}^{T} K_{f_1}^{(\ell)}(G_1(Z_{it}^0))K_{f_1}^{(\ell)}(G_1(Z_{il}^0)) \\
& -\frac{n^{-1/2}}{2\sigma^2}\mathrm{E}_0\Big[\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{1=l\neq t}^{T} K_{f_1}^{(\ell)}(G_1(Z_{it}^n))K_{f_1}^{(\ell)}(G_1(Z_{il}^n))\Big], \\
\mathrm{D}_2^{(n;\ell)} \;:=\; & \frac{n^{-1/2}}{2\sigma^2}\mathrm{E}_0\Big[\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{1=l\neq t}^{T} K_{f_1}^{(\ell)}(G_1(Z_{it}^n))K_{f_1}^{(\ell)}(G_1(Z_{il}^n))\Big],
\end{aligned}$$

$$\begin{aligned}
\mathrm{R}_1^{(n;\ell)} \;:=\; & \frac{n^{-1/2}}{2\sigma^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{1=l\neq t}^{T} K_{f_1}(G_1(Z_{it}^0))K_{f_1}(G_1(Z_{il}^0)) \\
& -\frac{n^{-1/2}}{2\sigma^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{1=l\neq t}^{T} K_{f_1}^{(\ell)}(G_1(Z_{it}^0))K_{f_1}^{(\ell)}(G_1(Z_{il}^0)),
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{R}_2^{(n;\ell)} \;:=\; & \frac{n^{-1/2}}{2\sigma^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{1=l\neq t}^{T} K_{f_1}(G_1(Z_{it}^n))K_{f_1}(G_1(Z_{il}^n)) \\
& -\frac{n^{-1/2}}{2\sigma^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{1=l\neq t}^{T} K_{f_1}^{(\ell)}(G_1(Z_{it}^n))K_{f_1}^{(\ell)}(G_1(Z_{il}^n)).
\end{aligned}$$

We prove that $\mathrm{C}^{(n)}(\boldsymbol{\theta}, \sigma^2) = o_{\mathrm{P}}(1)$ by establishing that $\mathrm{D}_1^{(n;\ell)}$ and $\mathrm{D}_2^{(n;\ell)}$ are $o_{\mathrm{P}}(1)$ under $\mathrm{P}^{(n)}_{\theta,\sigma^2,0;g_1}$, as $n \to \infty$, for fixed ℓ, and that $\mathrm{R}_1^{(n;\ell)}$ and $\mathrm{R}_2^{(n;\ell)}$ are $o_{\mathrm{P}}(1)$ under the same sequence of distributions, as $\ell \to \infty$, uniformly in n. For the sake of convenience, these three results are treated separately below (Lemmas 6.4, 6.5 and 6.6).

Decompose $\mathrm{D}_1^{(n;\ell)}$ into $\mathrm{D}_{1,1}^{(n;\ell)} + \mathrm{D}_{1,2}^{(n;\ell)} - \mathrm{E}_0[\mathrm{D}_{1,1}^{(n;\ell)}]$, where

$$\begin{aligned}\mathrm{D}_{1,1}^{(n;\ell)} \;:=\; & \frac{n^{-1/2}}{2\sigma^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{1=l\neq t}^{T} K_{f_1}^{(\ell)}(G_1(Z_{it}^n))K_{f_1}^{(\ell)}(G_1(Z_{il}^n)) \\ & -\frac{n^{-1/2}}{2\sigma^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{1=l\neq t}^{T} K_{f_1}^{(\ell)}(G_1(Z_{it}^0))K_{f_1}^{(\ell)}(G_1(Z_{il}^n))\end{aligned}$$

and

$$\begin{aligned}\mathrm{D}_{1,2}^{(n;\ell)} \;:=\; & \frac{n^{-1/2}}{2\sigma^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{1=l\neq t}^{T} K_{f_1}^{(\ell)}(G_1(Z_{it}^0))K_{f_1}^{(\ell)}(G_1(Z_{il}^n)) \\ & -\frac{n^{-1/2}}{2\sigma^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{1=l\neq t}^{T} K_{f_1}^{(\ell)}(G_1(Z_{it}^0))K_{f_1}^{(\ell)}(G_1(Z_{il}^0))\end{aligned}$$

(taking into account the independence between Z_{il}^0 and (Z_{it}^0, Z_{it}^n) under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,0;g_1}$). We then have the following.

Lemma 6.4 *For any fixed ℓ,*

(i) $\mathrm{E}_0\Big[\Big(\mathrm{D}_{1,1}^{(n;\ell)} - \mathrm{E}_0[\mathrm{D}_{1,1}^{(n;\ell)}]\Big)^2\Big] = o(1)$, *as $n\to\infty$;*

(ii) $\mathrm{E}_0\Big[\Big(\mathrm{D}_{1,2}^{(n;\ell)}\Big)^2\Big] = o(1)$, *as $n\to\infty$;*

(iii) $\mathrm{D}_1^{(n;\ell)} = o_{\mathrm{P}}(1)$, *as $n\to\infty$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,0;g_1}$.*

Lemma 6.5 *For any fixed ℓ, $\mathrm{D}_2^{(n;\ell)} = o(1)$, as $n\to\infty$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,0;g_1}$.*

Lemma 6.6 *As $\ell\to\infty$, uniformly in n,*

(a) $\mathrm{R}_1^{(n;\ell)}$ *is $o_{\mathrm{P}}(1)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,0;g_1}$,*

(b) $\mathrm{R}_2^{(n;\ell)}$ *is $o_{\mathrm{P}}(1)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,0;g_1}$ for n sufficiently large.*

Proof of Lemma 6.4 Let us begin with the second part of Lemma 6.4.
(ii) First note that $\mathrm{D}_{1,2}^{(n;\ell)} := (n^{-1/2}/2\sigma^2)\sum_{i=1}^{n}\mathrm{T}_i^{(n;\ell)}$, where, for $i=1,\ldots,n$,

$$\begin{aligned}\mathrm{T}_i^{(n;\ell)} \;:=\; & \sum_{t=1}^{T}\sum_{1=l\neq t}^{T} K_{f_1}^{(\ell)}(G_1(Z_{it}^0))\Big[K_{f_1}^{(\ell)}(G_1(Z_{il}^n)) - K_{f_1}^{(\ell)}(G_1(Z_{il}^0))\Big] \\ =\; & 2\sum_{t=1}^{T-1}\sum_{l=t+1}^{T} K_{f_1}^{(\ell)}(G_1(Z_{it}^0))\Big[K_{f_1}^{(\ell)}(G_1(Z_{il}^n)) - K_{f_1}^{(\ell)}(G_1(Z_{il}^0))\Big].\end{aligned}$$

are i.i.d. Using the independence, for $t \neq l$, between Z^0_{1t} and (Z^0_{1l}, Z^n_{1l}) and the boundedness of $K^{(\ell)}_{f_1}$, we have that

$$\begin{aligned}
\mathrm{E}_0\Big[\Big(\mathrm{D}^{(n;\ell)}_{1,2}\Big)^2\Big] &= \frac{n^{-1}}{4\sigma^4}\sum_{i,j=1}^{n}\mathrm{E}_0\Big[\mathrm{T}^{(n;\ell)}_i\mathrm{T}^{(n;\ell)}_j\Big] = \frac{n^{-1}}{4\sigma^4}\sum_{i=1}^{n}\mathrm{E}_0\Big[\Big(\mathrm{T}^{(n;\ell)}_i\Big)^2\Big] \\
&= \frac{n^{-1}}{\sigma^4}\sum_{i=1}^{n}\Big\{\sum_{t=1}^{T-1}\sum_{l=t+1}^{T}\mathrm{E}_0\Big[\Big(K^{(\ell)}_{f_1}(G_1(Z^0_{it}))\Big)^2\Big]\mathrm{E}_0\Big[\Big(K^{(\ell)}_{f_1}(G_1(Z^n_{il})) - K^{(\ell)}_{f_1}(G_1(Z^0_{il}))\Big)^2\Big] \\
&\quad + \sum_{t=1}^{T-1}\sum_{l\neq k=t+1}^{T}\mathrm{E}_0\Big[\Big(K^{(\ell)}_{f_1}(G_1(Z^0_{it}))\Big)^2\Big] \\
&\qquad \times\mathrm{E}_0\Big[\Big(K^{(\ell)}_{f_1}(G_1(Z^n_{il})) - K^{(\ell)}_{f_1}(G_1(Z^0_{il}))\Big)\Big(K^{(\ell)}_{f_1}(G_1(Z^n_{ik})) - K^{(\ell)}_{f_1}(G_1(Z^0_{ik}))\Big)\Big]\Big\} \\
&\leq \mathrm{C}_1 \max_{1\leq k\leq T}\Big\{\mathrm{E}_0\Big[\Big(K^{(\ell)}_{f_1}(G_1(Z^n_{ik})) - K^{(\ell)}_{f_1}(G_1(Z^0_{ik}))\Big)^2\Big]\Big\}.
\end{aligned}$$

Then, it only remains to be shown that $\mathrm{E}_0\Big[\Big(K^{(\ell)}_{f_1}(G_1(Z^n_{ik})) - K^{(\ell)}_{f_1}(G_1(Z^0_{ik}))\Big)^2\Big] = o(1)$, as $n \to \infty$, uniformly in k. Continuity of $K^{(\ell)}_{f_1} \circ G_1$ and the fact that $|Z^n_{il} - Z^0_{il}|$ is $o_{\mathrm{P}}(1)$ together imply that $K^{(\ell)}_{f_1}(G_1(Z^n_{il})) - K^{(\ell)}_{f_1}(G_1(Z^0_{il})) = o_{\mathrm{P}}(1)$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,0;g_1}$, as $n \to \infty$. Moreover, since $K^{(\ell)}_{f_1}$ is bounded, this convergence to zero also holds in quadratic mean.

(i) Letting $\mathrm{T}^{(n;\ell)}_i := \sum_{t=1}^{T-1}\sum_{l=t+1}^{T}\Big[K^{(\ell)}_{f_1}(G_1(Z^n_{it})) - K^{(\ell)}_{f_1}(G_1(Z^0_{it}))\Big]K^{(\ell)}_{f_1}(G_1(Z^n_{il}))$, we have

$$\mathrm{D}^{(n;\ell)}_{1,1} - \mathrm{E}_0[\mathrm{D}^{(n;\ell)}_{1,1}] = \frac{n^{-1/2}}{\sigma^2}\sum_{i=1}^{n}\Big[\mathrm{T}^{(n;\ell)}_i - \mathrm{E}_0[\mathrm{T}^{(n;\ell)}_i]\Big],$$

hence

$$\begin{aligned}
\mathrm{E}_0\Big[\Big(\mathrm{D}^{(n;\ell)}_{1,1} - \mathrm{E}_0[\mathrm{D}^{(n;\ell)}_{1,1}]\Big)^2\Big] &= \frac{n^{-1}}{\sigma^4}\sum_{i=1}^{n}\mathrm{E}_0\Big[\Big[\mathrm{T}^{(n;\ell)}_i - \mathrm{E}_0[\mathrm{T}^{(n;\ell)}_i]\Big]^2\Big] = \frac{n^{-1}}{\sigma^4}\sum_{i=1}^{n}\mathrm{Var}_0\Big[\mathrm{T}^{(n;\ell)}_i\Big] \\
&\leq \frac{n^{-1}}{\sigma^4}\sum_{i=1}^{n}\mathrm{E}_0\Big[\Big(\mathrm{T}^{(n;\ell)}_i\Big)^2\Big] = \frac{1}{\sigma^4}\mathrm{E}_0\Big[\Big(\mathrm{T}^{(n;\ell)}_1\Big)^2\Big].
\end{aligned}$$

It only remains to show that $\mathrm{E}_0\Big[(\mathrm{T}^{(n;\ell)}_1)^2\Big] = o(1)$, as $n \to \infty$:

$$\begin{aligned}
\mathrm{E}_0\Big[\Big(\mathrm{T}^{(n;\ell)}_1\Big)^2\Big] &= \sum_{t=1}^{T-1}\mathrm{E}_0\Big[\Big(K^{(\ell)}_{f_1}(G_1(Z^n_{1t})) - K^{(\ell)}_{f_1}(G_1(Z^0_{1t}))\Big)^2\Big]\mathrm{E}_0\Big[\Big(\sum_{l=t+1}^{T}K^{(\ell)}_{f_1}(G_1(Z^n_{1l}))\Big)^2\Big] \\
&\quad +2\sum_{t=2}^{T-1}\sum_{k=1}^{t-1}\sum_{l=t+1}^{T}\sum_{s=k+1}^{T}\mathrm{E}_0\Big[\Big(K^{(\ell)}_{f_1}(G_1(Z^n_{1t})) - K^{(\ell)}_{f_1}(G_1(Z^0_{1t}))\Big) \\
&\qquad \times\Big(K^{(\ell)}_{f_1}(G_1(Z^n_{1k})) - K^{(\ell)}_{f_1}(G_1(Z^0_{1k}))\Big)\,K^{(\ell)}_{f_1}(G_1(Z^n_{1l}))\,K^{(\ell)}_{f_1}(G_1(Z^n_{1s}))\Big] \\
&= \sum_{t=1}^{T-1}\mathrm{E}_0\Big[\Big(K^{(\ell)}_{f_1}(G_1(Z^n_{1t})) - K^{(\ell)}_{f_1}(G_1(Z^0_{1t}))\Big)^2\Big]\mathrm{E}_0\Big[\Big(\sum_{l=t+1}^{T}K^{(\ell)}_{f_1}(G_1(Z^n_{1l}))\Big)^2\Big] \\
&\quad +2\sum_{t=2}^{T-1}\sum_{k=1}^{t-1}\sum_{l=t+1}^{T}\mathrm{E}_0\Big[\Big(K^{(\ell)}_{f_1}(G_1(Z^n_{1t})) - K^{(\ell)}_{f_1}(G_1(Z^0_{1t}))\Big) \\
&\qquad \times\Big(K^{(\ell)}_{f_1}(G_1(Z^n_{1k})) - K^{(\ell)}_{f_1}(G_1(Z^0_{1k}))\Big)\,\Big(K^{(\ell)}_{f_1}(G_1(Z^n_{1l}))\Big)^2\Big]
\end{aligned}$$

$$+2\sum_{t=2}^{T-1}\sum_{k=1}^{t-1}\sum_{l=t+1}^{T}\mathrm{E}_0\Big[\Big(K_{f_1}^{(\ell)}(G_1(Z_{1t}^n))-K_{f_1}^{(\ell)}(G_1(Z_{1t}^0))\Big)K_{f_1}^{(\ell)}(G_1(Z_{1t}^n))$$
$$\times\Big(K_{f_1}^{(\ell)}(G_1(Z_{1k}^n))-K_{f_1}^{(\ell)}(G_1(Z_{1k}^0))\Big)K_{f_1}^{(\ell)}(G_1(Z_{1l}^n))\Big]$$
$$+2\sum_{t=2}^{T-1}\sum_{k=1}^{t-1}\sum_{l=t+1}^{T}\sum_{s\neq l,s,t=k+1}^{T}\mathrm{E}_0\Big[\Big(K_{f_1}^{(\ell)}(G_1(Z_{1t}^n))-K_{f_1}^{(\ell)}(G_1(Z_{1t}^0))\Big)$$
$$\times\Big(K_{f_1}^{(\ell)}(G_1(Z_{1k}^n))-K_{f_1}^{(\ell)}(G_1(Z_{1k}^0))\Big)K_{f_1}^{(\ell)}(G_1(Z_{1l}^n))\,K_{f_1}^{(\ell)}(G_1(Z_{1s}^n))\Big]$$
$$\leq\ \mathrm{C}_2\max_{1\leq k\leq T}\Big\{\mathrm{E}_0\Big[\Big(K_{f_1}^{(\ell)}(G_1(Z_{1k}^n))-K_{f_1}^{(\ell)}(G_1(Z_{1k}^0))\Big)^2\Big]\Big\}=o(1).$$

(iii) trivially follows from (i)-(ii) and the fact that convergence in quadratic mean implies convergence in probability. □

Proof of Lemma 6.5. Letting $\mathrm{B}_1^{(n;\ell)} := \frac{n^{-1/2}}{2\sigma^2}\sum_{i=1}^n\sum_{t=1}^T\sum_{1=l\neq t}^T K_{f_1}^{(\ell)}(G_1(Z_{it}^0))K_{f_1}^{(\ell)}(G_1(Z_{il}^0))$ and $\mathrm{B}_2^{(n;\ell)} := \frac{n^{-1/2}}{2\sigma^2}\sum_{i=1}^n\sum_{t=1}^T\sum_{1=l\neq t}^T K_{f_1}^{(\ell)}(G_1(Z_{it}^n))K_{f_1}^{(\ell)}(G_1(Z_{il}^n))$, we have, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,0;g_1}$,

$$\mathrm{B}_1^{(n;\ell)}\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,\frac{T(T-1)}{2\sigma^4}\Big(\mathrm{E}[(K_{f_1}^{(\ell)}(U))^2]\Big)^2\right)$$

as $n\to\infty$, where U stands for a random variable uniformly distributed over $(0,1)$, and

$$\mathrm{B}_2^{(n;\ell)}\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,\frac{T(T-1)}{2\sigma^4}\Big(\mathrm{E}[(K_{f_1}^{(\ell)}(U))^2]\Big)^2\right). \tag{6.14}$$

Since $\mathrm{D}_1^{(n;\ell)}=\mathrm{B}_2^{(n;\ell)}-\mathrm{B}_1^{(n;\ell)}-\mathrm{E}_0[\mathrm{B}_2^{(n;\ell)}]=o_{\mathrm{P}}(1)$, it follows that, still under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,0;g_1}$,

$$\mathrm{B}_2^{(n;\ell)}-\mathrm{E}_0[\mathrm{B}_2^{(n;\ell)}]\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,\frac{T(T-1)}{2\sigma^4}\Big(\mathrm{E}[(K_{f_1}^{(\ell)}(U))^2]\Big)^2\right) \tag{6.15}$$

as $n\to\infty$. From (6.14) and (6.15), it follows that $\mathrm{D}_2^{(n;\ell)}=\mathrm{E}_0[\mathrm{B}_2^{(n;\ell)}]$ is $o(1)$ as $n\to\infty$. □

Proof of Lemma 6.6 (a) We have that

$$\mathrm{E}_0[(\mathrm{R}_1^{(n;\ell)})^2]$$
$$=\frac{n^{-1}}{\sigma^4}\sum_{i=1}^n\mathrm{E}_0\Big[\Big(\sum_{t=1}^{T-1}\sum_{l=t+1}^{T}\Big[K_{f_1}(G_1(Z_{it}^0))K_{f_1}(G_1(Z_{il}^0))-K_{f_1}^{(\ell)}(G_1(Z_{it}^0))K_{f_1}^{(\ell)}(G_1(Z_{il}^0))\Big]\Big)^2\Big]$$
$$=\frac{n^{-1}}{\sigma^4}\sum_{i=1}^n\sum_{t=1}^{T-1}\sum_{l=t+1}^{T}\mathrm{E}_0\Big[\Big(K_{f_1}(G_1(Z_{it}^0))K_{f_1}(G_1(Z_{il}^0))-K_{f_1}^{(\ell)}(G_1(Z_{it}^0))K_{f_1}^{(\ell)}(G_1(Z_{il}^0))\Big)^2\Big]$$
$$+\frac{n^{-1}}{\sigma^4}\sum_{i=1}^n\sum_{t=1}^{T-1}\sum_{t+1\leq l\neq k\leq T}\mathrm{E}_0\Big[\Big(K_{f_1}(G_1(Z_{it}^0))K_{f_1}(G_1(Z_{il}^0))-K_{f_1}^{(\ell)}(G_1(Z_{it}^0))K_{f_1}^{(\ell)}(G_1(Z_{il}^0))\Big)$$
$$\times\Big(K_{f_1}(G_1(Z_{it}^0))K_{f_1}(G_1(Z_{ik}^0))-K_{f_1}^{(\ell)}(G_1(Z_{it}^0))K_{f_1}^{(\ell)}(G_1(Z_{ik}^0))\Big)\Big]$$
$$+\frac{2n^{-1}}{\sigma^4}\sum_{i=1}^n\sum_{t=2}^{T-1}\sum_{k=1}^{t-1}\sum_{l=t+1}^{T}\sum_{s=k+1}^{T}\mathrm{E}_0\Big[\Big(K_{f_1}(G_1(Z_{it}^0))K_{f_1}(G_1(Z_{il}^0))-K_{f_1}^{(\ell)}(G_1(Z_{it}^0))K_{f_1}^{(\ell)}(G_1(Z_{il}^0))\Big)$$
$$\times\Big(K_{f_1}(G_1(Z_{ik}^0))K_{f_1}(G_1(Z_{is}^0))-K_{f_1}^{(\ell)}(G_1(Z_{ik}^0))K_{f_1}^{(\ell)}(G_1(Z_{is}^0))\Big)\Big]$$
$$=:\ A_1+A_2+A_3,\ \text{ say.}$$

Using the fact that $\mathrm{E}_0[K_{f_1}^{(\ell)}(G_1(Z_{it}^0))] = o(1)$ as $\ell \to \infty$, uniformly in n and for all t and i, it follows that A_2 and A_3 are $o(1)$ as $\ell \to \infty$, uniformly in n and it only remains to show that A_1 is $o(1)$ as $\ell \to \infty$, uniformly in n. Now, we have that

$$
\begin{aligned}
A_1 &= \frac{n^{-1}}{\sigma^4} \sum_{i=1}^{n} \sum_{t=1}^{T-1} \sum_{l=t+1}^{T} \mathrm{E}_0\Big[\Big(K_{f_1}(G_1(Z_{it}^0))K_{f_1}(G_1(Z_{il}^0)) - K_{f_1}^{(\ell)}(G_1(Z_{it}^0))K_{f_1}^{(\ell)}(G_1(Z_{il}^0))\Big)^2\Big] \\
&= \frac{n^{-1}}{\sigma^4} \sum_{i=1}^{n} \sum_{t=1}^{T-1} \sum_{l=t+1}^{T} \int_0^1 \int_0^1 \Big(K_{f_1}(u)K_{f_1}(v) - K_{f_1}^{(\ell)}(u)K_{f_1}^{(\ell)}(v)\Big)^2 du\, dv \\
&= \frac{T(T-1)}{2\sigma^4} \int_0^1 \int_0^1 \Big(K_{f_1}(u)K_{f_1}(v) - K_{f_1}^{(\ell)}(u)K_{f_1}^{(\ell)}(v)\Big)^2 du\, dv. \qquad (6.16)
\end{aligned}
$$

Since $K_{f_1}^{(\ell)}(u)K_{f_1}^{(\ell)}(v)$ converges to $K_{f_1}(u)K_{f_1}(v)$ for all $(u,v) \in (0,1)\times(0,1)$, and since $|K_{f_1}^{(\ell)}(u)|$ is bounded by $|K_{f_1}(u)|$ for all $\ell \geq L$, the integrand in (6.16) is bounded (uniformly in ℓ) by $4|K_{f_1}(u)|^2|K_{f_1}(v)|^2$, which is integrable on $(0,1)\times(0,1)$. The Lebesgue dominated convergence theorem implies that $A_1 = o(1)$ as $\ell \to \infty$, uniformly in n.

(b) The claim here is the same as in (a), except that Z_{it}^n replaces Z_{it}^0. Accordingly, (b) holds under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}^{(n)},\sigma_n^2,0;g_1}$. That is also holds under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,0;g_1}$ follows from Lemma 3.5 in Jurečkova (1969). □

References

[1] Akharif, A., and Hallin, M. (2003). Efficient detection of random coefficients in autoregressive models. *The Annals of Statistics* **31**, 675-704.

[2] Arellano, M. (2003). *Panel Data Econometrics*. Oxford University Press.

[3] Azzalini, A., and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *Journal of the Royal Statistical Society Series B* **65**, 367-389.

[4] Baltagi, B.H. (2008). *Econometric Analysis of Panel Data*, 4rth edition. J. Wiley, Chichester.

[5] Bickel, P.J., Klaassen, C.A.J., Ritov, Y., and Wellner, J.A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models. Johns Hopkins Univ. Press*, Baltimore, MD.

[6] Breusch, T.S., and Pagan, A.R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies* **47**, 239-253.

[7] Chesher, A.D. (1984). Testing for neglected heterogeneity. *Econometrica* **52**, 865-872.

[8] Häggström, E., and Laitila, T. (2002). Test of random subject effects in heteroskedastic linear models. *Biometrical Journal* **44**, 825-834.

[9] Hájek, J., and Šidák, Z. (1967). *Theory of Rank Tests*. Academic Press, New York.

[10] Hájek, J., Šidák, Z., and Sen, P.K. (1999). *Theory of Rank Tests*. 2nd ed. Academic Press, New York.

[11] Hallin, M., Ingenbleek, J.-F., and Puri, M.L. (1985). Linear serial rank tests for randomness against ARMA alternatives. *The Annals of Statistics* **13**, 1156-1181.

[12] Hallin, M., and Werker, B. (2003). Semiparametric efficiency, distribution-freeness, and invariance. *Bernoulli* **9**, 137-165.

[13] Hamerle, A. (1990). On a simple test for neglected heterogeneity in panel studies. *Biometrics* **46**, 193-199.

[14] Honda, Y. (1985). Testing the error components model with non-Normal disturbances. *Review of Economic Studies* **52**, 681-690.

[15] Hsiao, C. (2003). *Analysis of Panel Data*, 2nd Edition. Cambridge University Press.

[16] Jacqmin-Gadda, H., and Commenges, D. (1995). Tests of homogeneity for generalized linear models. *Journal of the American Statistical Association* **90**, 1237-1246.

[17] Jurečková, J. (1969). Asymptotic linearity of a rank statistic in regression parameter. *The Annals of Mathematical Statistics* **40**, 1889-1900.

[18] King, M.L., and Evans, M.A. (1986). Testing for block effects in regression models based on survey data. *Journal of the American Statistical Association* **81**, 677-679.

[19] Kreiss, J.P. (1987). On adaptive estimation in stationary ARMA processes. *Annals of Statistics* **15**, 112-133.

[20] Le Cam, L.M. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag, New York.

[21] Le Cam, L.M. and Yang, G. L. (2000). *Asymptotics in Statistics*, 2nd edition. Springer-Verlag, New York.

[22] Lin, X. (1997). Variance component testing in generalised linear models with random effects. *Biometrika* **84**, 309-326.

[23] Moulton, B.R. (1986). Random group effects and the precision of regression estimates. *Journal fo Econometrics* **32**, 385-397.

[24] Moulton, B.R., and Randolph, W.C. (1989). Alternative tests of the error components model. *Econometrica* **57**, 685-693.

[25] Nerlove, M. (1971). Further evidence on the estimation of dynamic economic relations from a time-series of cross-sections. *Econometrica* **39**, 359-382.

[26] Orme, C.D. (1993). A comment on “A simple test for neglected heterogeneity in panel studies”. *Biometrics* **49**, 665-667.

[27] Orme, C.D., and Yamagata, T. (2006). The asymptotic distribution of the F-test statistic for individual effects. *Econometrics Journal* **9**, 404-422.

[28] Scott, A.J., and Holt, D. (1982). The effects of Two-stage sampling on ordinary least squares methods. *Journal of the American Statistical Association* **77**, 848-854.

[29] Swensen, A.R. (1985), The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *Journal of Multivariate Analysis* **16**, 54-70.

[30] van der Vaart, A.W. (1988). *Asymptotic Statistic*. Cambridge University Press, Cambridge.

[31] Wooldridge, J.M. (2002), *Econometric Analysis of Cross-Section and Panel Data*. MIT Press, Cambridge, Mass.