arXiv:submit/7407362 [econ.EM] 25 Mar 2026

Refined Cluster Robust Inference*

Bulat Gafarov† Takuya Ura‡

March 25, 2026

Abstract

It has become standard for empirical studies to conduct inference robust to cluster dependence and heterogeneity. With a small number of clusters, the normal approximation for the t-statistics of regression coefficients may be poor. This paper tackles this problem using a critical value based on the conditional Cramér-Edgeworth expansion for the t-statistics. Our approach guarantees third-order refinement, regardless of whether a regressor is discrete or not, and, unlike the cluster pairs bootstrap, avoids resampling data. Simulations show that our proposal can make a difference in size control with as few as 10 clusters.

Keywords: Cluster robust inference, Cramér-Edgeworth expansion, Asymptotic refinement

JEL Classification: C12, C21

*We would like to thank A. Colin Cameron for very helpful comments. Ura acknowledges funding from A. Colin Cameron Associate Professor Research Award.

†Department of Agricultural and Resource Economics, University of California, Davis. Email: bgafarov@ucdavis.edu

‡Department of Economics, University of California, Davis. Email: takura@ucdavis.edu

1 Introduction

Cluster robust inference has become a standard practice in applied microeconometrics. Robustness to arbitrary dependence within a cluster often comes at the cost of a reduction in the effective sample size. As a result, despite a large overall sample size, the researchers have to account for the non-Gaussian distribution of the t-statistic for significance tests as was the case in classical statistics Student (1908). In particular, it is well known that the standard normality-based inference method may lead to over-rejection when the number of clusters is small (Cameron and Miller, 2015, 2025; MacKinnon, Nielsen, and Webb, 2023). Instead of assuming Gaussian regression errors, one can use only mild moment restrictions and derive corrected critical values using higher order asymptotic expansions of Edgeworth (1905) and Cramér (1928).

In this paper, we propose a new analytical correction to critical values based on inverting the Cramér-Edgeworth expansion of the t-statistic null distribution. The resulting inference method is third-order asymptotically accurate and robust against heterogeneous cluster dependence. We first derive the Cramér-Edgeworth term up to the second order and then adjust the critical value based on the estimated Cramér-Edgeworth term.

We develop the approach of Hall (1983) in the setup of linear regression with heterogeneous clusters.[1] As special sub-cases, our approach also nests inference on the sample mean of non-identically distributed data and regression coefficients in cross-sectional regression. The resulting inference is third-order asymptotically accurate in the sense that the actual test size differs from the prespecified significance level α only by $o(G^{-1})$ with G clusters.

Our simulation studies support this theoretical size control.

A cluster pairs bootstrap might be a popular choice for asymptotic refinement, but our proposed method has a few advantages over the cluster pairs bootstrap.[2] First and most importantly, existing simulation studies show that the cluster pairs bootstrap does not perform well with a small number of clusters. For example, Cameron, Gelbach, and Miller (2008) use a simulation design based on Bertrand, Duflo, and Mullainathan (2004) and explain that the poor performance of the cluster pairs bootstrap is due to the fact that the resampled values of the Gram matrix $X'X$ are nearly singular. Our approach of analytically inverting the Cramér-Edgeworth expansion avoids this problem by not resampling $X'X$ while achieving asymptotic refinement. Second, the cluster pairs bootstrap uses independence across $X_1, \ldots, X_G$, while our expansion does not require it. We allow $X_1, \ldots, X_G$ to be correlated, e.g., through adaptive randomization. Third, the standard proof for the cluster pairs bootstrap's asymptotic refinement (e.g., Liu, 1988; Hall, 2013) excludes discrete regressors. To apply the results from Hall (2013, Ch.5) to the regression framework, we need to impose the Cramér's condition on the regressors, but the Cramér's condition fails for discrete random variables (Bhattacharya and Rao, 1976, p.207).[3] Last, our proposed critical value has a closed-form

[1] In this paper, we consider a two-sided alternative hypothesis and the second-order Cramér-Edgeworth expansion. As a result, we do not need to apply the Cramér-Edgeworth expansion of the estimated coefficient recursively, as in Hall (1983).

[2] The residual bootstrap cannot be applied for cluster robust inference when the sample sizes vary across clusters.

[3] We could not find a sufficient condition for the cluster pairs bootstrap's asymptotic refinement that allows for

expression, and for this reason, its computation is much faster than resampling.

A number of other methods have been proposed for the construction of standard errors and confidence intervals. Our proposal is unique in the following sense: it demonstrates good finite-sample performance in the simulation designs based on Bertrand et al. (2004), while simultaneously achieving third-order asymptotic refinement. For example, Cameron et al. (2008) propose the wild cluster bootstrap and demonstrate its good finite-sample properties in simulations. Djogbenou, MacKinnon, and Nielsen (2019) provide the asymptotic size control of the wild cluster bootstrap, and Canay, Santos, and Shaikh (2021) show the size control of the wild cluster bootstrap even with a fixed number of the clusters as long as there are a large number of observations per cluster. However, Theorem 5.2 of Djogbenou et al. (2019) shows that the wild cluster bootstrap does not achieve asymptotic refinement when the score has non-zero skewness. Our simulation results in Section 3.2 confirm this result of Djogbenou et al. (2019), and our proposed method demonstrates better size control than the wild cluster bootstrap in such cases.

This paper is also related to the use of the t-distribution for the critical value and an adjustment of degrees of freedom in the t-distribution (McCaffrey and Bell, 2003; Bester, Conley, and Hansen, 2011; Imbens and Kolesar, 2016; Young, 2016; Hansen, 2021, 2025). These papers assume normal and homoscedastic errors. On the other hand, our approach does not rely on normal errors or a specific covariance structure for the error terms within a cluster.

The paper proceeds as follows. Section 2 formally introduces the regression model with clustered errors and the new critical value. Section 3 presents Monte Carlo simulations for the proposed critical value and existing ones. Section 4 concludes. The appendix collects all the proofs and additional results.

2 Inverting a Cramér-Edgeworth Expansion with Clustered Errors

We have the dataset of $\{(Y_{ig}, X_{ig}) : i = 1, \ldots, N_g,\ g = 1, \ldots, G\}$ and consider the regression model $Y_{ig} = X_{ig}'\beta + u_{ig}$ with $E[u_{ig} \mid X_{1g}, \ldots, X_{N_g g}] = 0$ and $\dim(X_{ig}) = k$. We assume the observations $\{(Y_{ig}, X_{ig}) : i = 1, \ldots, N_g\}$ are independent across $g = 1, \ldots, G$. To emphasize the fact that we use the independence across g, we use the matrix notation with $Y_g = (Y_{1g}, \ldots, Y_{N_g g})'$ and $X_g = (X_{1g}', \ldots, X_{N_g g}')'$, and write the regression model succinctly as

$$Y_g = X_g\beta + u_g \text{ with } E[u_g \mid X_g] = 0.$$

For a fixed vector λ and a hypothesized value c_0 for $\lambda'\beta$, we consider the hypothesis testing problem of

$$H_0 : \lambda'\beta = c_0 \text{ vs } H_1 : \lambda'\beta \neq c_0$$

discrete regressors and non-identical distributions. A weaker version of the Cramér's condition has been proposed, e.g., in Bai and Rao (1991), but the cluster pairs bootstrap's asymptotic refinement without the classical Cramér's condition is beyond the scope of this paper.

with the significance level $\alpha \in (0,1)$. The OLS estimator for β is

$$\hat{\beta} = \left(\frac{1}{G} \sum_{g=1}^{G} X_g' X_g \right)^{-1} \left(\frac{1}{G} \sum_{g=1}^{G} X_g' Y_g \right).$$

Consider an asymptotic variance estimator for $\lambda'\hat{\beta}$ defined by

$$\hat{\sigma}^2 = \frac{1}{G} \sum_{g=1}^{G} (\lambda' \Pi X_g' \hat{u}_g)^2 \text{ with } \Pi = \left(\frac{1}{G} \sum_{g=1}^{G} X_g' X_g \right)^{-1} \text{ and } \hat{u}_g = Y_g - X_g \hat{\beta}.$$

Define the t-statistic by

$$t = \sqrt{G} \frac{\lambda' \hat{\beta} - c_0}{\hat{\sigma}}.$$

From now on, we estimate the null distribution for the above t-statistic t and construct a critical value for it. We treat the covariates $\mathbf{X} = \{X_g\}_{g=1}^{\infty}$ as fixed, so we investigate $Pr(|t| \leq z \mid \mathbf{X} = \mathbf{x})$ for a given sequence of constants $\mathbf{x} = \{x_g\}_{g=1}^{\infty}$.

We consider the numerator and denominator of

$$t = \sqrt{G} \frac{(\lambda' \hat{\beta} - c_0)/\sigma}{\hat{\sigma}/\sigma}$$

under the null hypothesis H_0, where σ^2 is the asymptotic variance for $\lambda'\hat{\beta}$ defined by

$$\sigma^2 = \frac{1}{G} \sum_{g=1}^{G} \sigma_g^2 \text{ with } \sigma_g^2 = E\left[(\lambda' \Pi X_g' u_g)^2 \mid \mathbf{X} = \mathbf{x} \right].$$

The numerator has the linear representation of

$$(\lambda' \hat{\beta} - c_0)/\sigma = \frac{1}{G} \sum_{g=1}^{G} \omega_{1g} \text{ with } \omega_{1g} = \sigma^{-1} \lambda' \Pi X_g' u_g.$$

The square of the denominator $\hat{\sigma}^2/\sigma^2$ has the following quadratic representation. The proof is given in Section A.1.

Lemma 2.1.

$$\hat{\sigma}^2/\sigma^2 = 1 - \left(\frac{1}{G} \sum_{g=1}^{G} \omega_{2g} \right)' \Gamma \left(\frac{1}{G} \sum_{g=1}^{G} \omega_{2g} \right) + \frac{1}{G} \sum_{g=1}^{G} \omega_{3g}$$

where

$$\Gamma = \begin{pmatrix} -\frac{1}{G} \sum_{g=1}^{G} X_g' X_g \Pi' \lambda \lambda' \Pi X_g' X_g & I_k \\ I_k & O \end{pmatrix}$$

and

$$\omega_{2g} = \sigma^{-1} \begin{pmatrix} I_k \\ X_g' X_g \Pi' \lambda \lambda' \end{pmatrix} \Pi X_g' u_g,$$
$$\omega_{3g} = \sigma^{-2}((\lambda' \Pi X_g' u_g)^2 - \sigma_g^2).$$

To approximate the null distribution for the t-statistic, we approximate the distribution of $\frac{1}{\sqrt{G}} \sum_{g=1}^{G} (\omega_{1g}, \omega_{2g}', \omega_{3g})'$ up to $o(G^{-1})$. For this purpose, we use the following moments of ω_{1g} and ω_{2g}:

$$\mu_{1,2} = \frac{1}{G} \sum_g E\left[\omega_{1g}\omega_{2g} \mid \mathbf{X} = \mathbf{x}\right],$$
$$\mu_{2,2} = \frac{1}{G} \sum_g E\left[\omega_{2g}'\Gamma\omega_{2g} \mid \mathbf{X} = \mathbf{x}\right],$$
$$\mu_{1,1,1} = \frac{1}{G} \sum_g E\left[\omega_{1g}^3 \mid \mathbf{X} = \mathbf{x}\right],$$
$$\mu_{1,1,1,1} = \frac{1}{G} \sum_g E\left[\omega_{1g}^4 \mid \mathbf{X} = \mathbf{x}\right]$$

We assume these moments are bounded.

Assumption 1. *$\mu_{1,2}$, $\mu_{2,2}$, $\mu_{1,1,1}$, and $\mu_{1,1,1,1}$ are bounded uniformly in G.*[4]

The (second-order) Cramér-Edgeworth expansion for the t-statistic's null distribution is expressed under the following assumption, for which we provide a sufficient condition in Section 2.1.

Assumption 2. *Under H_0, there is a neighborhood $\mathcal{N}$ of $\Phi^{-1}(1-\alpha/2)$ such that*

$$\sup_{z \in \mathcal{N}} |Pr(|t| \leq z \mid \mathbf{X} = \mathbf{x}) - (2\Phi(z) - 1 + 2G^{-1}q_2(z)\phi(z))| = o(G^{-1}),$$

where He_r are the r-th order Hermite polynomials and

$$q_2(z) = -\left(\frac{1}{2}(k_2 + k_1^2)He_1(z) + \frac{1}{24}(k_4 + 4k_1k_3)He_3(z) + \frac{1}{72}k_3^2 He_5(z)\right).$$

The parameters $k_1, \ldots, k_4$ are defined as

$$k_1 = \nu_1$$
$$k_2 = \nu_2 - \nu_1^2$$
$$k_3 = \nu_3 - 3\nu_1$$
$$k_4 = \nu_4 - 4\nu_1\nu_3 - 6\nu_2 + 12\nu_1^2$$

[4] All the results in this paper only require the assumptions hold for sufficiently large G. To simplify the exposition, we impose the assumptions for every value of G.

where

$$\begin{aligned}
\nu_1 &= -\frac{\mu_{1,1,1}}{2} \\
\nu_2 &= 2\mu_{1,1,1}^2 + (\mu_{2,2} + 2\mu_{1,2}'\Gamma\mu_{1,2}) \\
\nu_3 &= -\frac{7}{2}\mu_{1,1,1} \\
\nu_4 &= -2\mu_{1,1,1,1} + 28\mu_{1,1,1}^2 + 6\mu_{2,2} + 24\mu_{1,2}'\mu_{1,2}.
\end{aligned}$$

The function $q_2(z)$ in the (second-order) Cramér-Edgeworth expansion is unknown since we do not know the population objects of $\mu_{1,2}'\mu_{1,2}, \mu_{2,2}, \mu_{1,1,1}, \mu_{1,1,1,1}$. We can estimate them using their sample analogs:

$$\begin{aligned}
\hat{\mu}_{1,2}'\Gamma\hat{\mu}_{1,2} &= \left(\frac{1}{G}\sum_g \hat{\omega}_{1g_1}\hat{\omega}_{2g_1}\right)' \Gamma \left(\frac{1}{G}\sum_g \hat{\omega}_{1g}\hat{\omega}_{2g}\right) \\
\hat{\mu}_{2,2} &= \frac{1}{G}\sum_g \hat{\omega}_{2g}'\Gamma\hat{\omega}_{2g}, \\
\hat{\mu}_{1,1,1} &= \frac{1}{G}\sum_g \hat{\omega}_{1g}^3, \\
\hat{\mu}_{1,1,1,1} &= \frac{1}{G}\sum_g \hat{\omega}_{1g}^4,
\end{aligned}$$

where

$$\begin{aligned}
\hat{\omega}_{1g} &= \hat{\sigma}^{-1}\lambda'\Pi X_g'\hat{u}_g, \\
\hat{\omega}_{2g} &= \hat{\sigma}^{-1}\begin{pmatrix} I_k \\ X_g'X_g\Pi'\lambda\lambda' \end{pmatrix}\Pi X_g'\hat{u}_g.
\end{aligned}$$

We can construct the estimator $\hat{q}_2(z)$ for $q_2(z)$ using these sample analogs. Our proposed critical value is

$$\hat{cv} = \Phi^{-1}(1-\alpha/2) - G^{-1}\hat{q}_2(\Phi^{-1}(1-\alpha/2))$$

and the resulting confidence interval for $\lambda'\beta$ is $\lambda'\hat{\beta} \pm \hat{cv}\sqrt{\hat{\sigma}/G}$.

We impose the following conditions on moments.

Assumption 3. *(i) σ^2 and the minimum eigenvalue of $\frac{1}{G}\sum_{g=1}^G x_g'x_g$ are bounded away from zero uniformly in G. (ii) $\frac{1}{G}\sum_{g=1}^G \left\|x_g'x_g\right\|^4$ and $\frac{1}{G}\sum_{g=1}^G E[\left\|X_g'u_g\right\|^j \mid \mathbf{X} = \mathbf{x}]$ are bounded uniformly in G for every positive integer j.*

The second condition makes the characteristic function of $\frac{1}{G}\sum_{g=1}^G \omega_g$ infinitely differentiable and simplifies the proofs. We may weaken it to the bounded moment condition up to a certain order by truncating ω_g (cf., Hall, 2013, p.256; Bhattacharya and Ghosh, 1978, p.446). The bounded higher-order moments are crucial for our inference because we estimate the population objects of

$\mu_{1,2}, \mu_{2,2}, \mu_{1,1,1}, \mu_{1,1,1,1}$. For example, $\mu_{1,1,1,1}$ is the average fourth moment of ω_{1g}. In the proof, we use the fourth moment of its estimator and thus require bounded 16th moments.

In Theorem 1 below, we show the size control for the proposed critical value. Note that even if Assumption 2 fails, we can still achieve size control with an asymptotic approximation error of the standard rate $O(G^{-1})$, as long as asymptotic normality holds. This point resembles the efficiency gain of the feasible generalized least squares estimation (Cameron and Miller, 2025, Section 3.5).

Theorem 1. *$Pr(|t| \leq \hat{cv} \mid \mathbf{X} = \mathbf{x}) = 1 - \alpha + o(G^{-1})$ under H_0 and Assumption 1-3.*

The proof is provided in Section A.2.

Remark 1. *Our proposed confidence interval does not rely on a particular choice of the asymptotic variance estimator in the following sense. Suppose we use another asymptotic variance estimator for $\lambda'\hat{\beta}$ denoted by $\tilde{\sigma}^2$. Then the t-statistic becomes*

$$\sqrt{G}\frac{\lambda'\hat{\beta} - c_0}{\tilde{\sigma}} = \frac{\hat{\sigma}}{\tilde{\sigma}}t.$$

We can accordingly change the critical value as $(\hat{\sigma}/\tilde{\sigma})\hat{cv}$, but the resulting confidence interval is still $\lambda'\hat{\beta} \pm \hat{cv}\sqrt{\hat{\sigma}/G}$.

Remark 2. *In this paper, we treated $\mathbf{X}$ as fixed, so we do not apply the Cramér-Edgeworth expansion to $\frac{1}{G}\sum_{g=1}^{G} X_g'X_g$. It is different from Djogbenou et al. (2019), who consider the regressor $\mathbf{X}$ as random when they derive the Cramér-Edgeworth expansion for the t-statistic t.*[5] *First, our expansion does not require the independence of X_g across g. As a consequence, we allow $X_1, \ldots, X_G$ to be generated from an adaptive randomization. Second, the assumption in Djogbenou et al. (2019) for the Cramér-Edgeworth expansion excludes empirically relevant cases such as binary regressors. With the random regressor, they apply the Cramér-Edgeworth expansion to the term $\frac{1}{G}\sum_{g=1}^{G} X_g'X_g$ and assume the Cramér's condition on $X_g'X_g$. This condition fails if X_g includes discrete variables (Bhattacharya and Rao, 1976, p.207).*[6] *In this paper, we do not require the Cramér's condition on $X_g'X_g$. Third, the resulting Cramér-Edgeworth expansion has fewer unknown parameters to estimate. Namely, we do not need to estimate the moments of $X_g'X_g$.*

Remark 3. *The above Cramér-Edgeworth expansion for the t-statistic in Assumption 2 is different from the one for the (standardized) sample mean $\sigma^{-1}\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\lambda'\Pi X_g'u_g$ since we estimate the average variance σ^2. One of the qualitative differences is that we can estimate the above $q_2(z)$ using*

[5] They do not use the Cramér-Edgeworth expansion to construct an asymptotically refined inference, but instead use it to show the wild bootstrap cannot achieve the asymptotic refinement of $o(G^{-1})$. As a result, this remark is not relevant for their analysis on the wild bootstrap because their conclusion relies on the fact that the wild bootstrap cannot replicate the first four cumulants of the Cramér-Edgeworth expansion.

[6] Consider the case where X_{ig} includes a binary variable D_g. The matrix $X_g'X_g$ includes $N_gD_g^2$, which takes the two values of N_g and 0. By choosing the vector t such that $t'\text{vech}(X_g'X_g) = \|t\|N_gD_g^2$, we have

$$E[\exp(it'\text{vech}(X_g'X_g))] = Pr(D_g = 1)\exp(i\|t\|N_g) + (1 - Pr(D_g = 1)) = 1 \text{ if } \|t\|N_g \text{ is a multiple of } 2\pi.$$

It implies that this Cramér's condition fails.

the sample analogs for the coefficients, while we cannot estimate the Cramér-Edgeworth expansion for $\sigma^{-1}\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\lambda'\Pi X_g'u_g$ *without additional assumptions. In this sense, even if we know the value of* σ^2*, we need to use the estimated value of* σ^2 *for the inference strategy proposed in this paper.*[7]

2.1 Cramér-Edgeworth Expansion in Assumption 2

In this subsection, we provide a sufficient condition for the Cramér-Edgeworth expansion in Assumption 2.

Assumption 4. *There are random vectors* $\eta_1,\dots,\eta_G$*, and matrices* $\mathbf{M}_1$ *and* $\mathbf{M}_2$ *such that* $\omega_g = \mathbf{M}_1\eta_g$ *and* $\eta_g = \mathbf{M}_2\omega_g$ *and that the minimum eigenvalue of* $\mathbf{V}_G = Var(\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\eta_g \mid \mathbf{X}=\mathbf{x})$ *is bounded away from zero uniformly in* G*.*

This assumption removes the redundant or duplicated elements from ω_g. This removal is necessary to normalize the random variable $\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\eta_g$ by using the matrix square root of the variance matrix $\mathbf{V}_G$.

Assumption 5. *There are positive numbers* C, R, b *with* $b<2$ *such that* $\frac{1}{G}\sum_{g=1}^{G}|E\left[\exp\left(it'\eta_g\right)\mid \mathbf{X}=\mathbf{x}\right]| \leq 1-C\|t\|^{-b}$ *for every* t *with* $\|t\|>R$ *and for sufficiently large* G*.*

The assumption is the mean weak Cramér's condition proposed in Angst and Poly (2017). We provide a sufficient condition for Assumption 5.

Theorem 2. *Suppose* u_g *has an absolutely continuous distribution given* $\mathbf{X}=\mathbf{x}$ *(with respect to the* N_g*-dimensional Lebesgue measure), and there are positive numbers* R, c_1, c_2 *such that* $|\mathbf{B}(t)| \geq c_2 G$ *for every* t *with* $\|t\|=1$ *and sufficiently large* G*, where* $\mathbf{B}(t) = \{g : \pi^2/(16R^2) \leq Var(t'\eta_g \mid \mathbf{X}=\mathbf{x}) \leq c_1 \mathrm{med}(f_{t'\eta_g|\mathbf{X}=\mathbf{x}}(t'\eta_g)\mid \mathbf{X}=\mathbf{x})^{-2}\}$.[8] *Then Assumption 5 holds.*

[7] The Cramér-Edgeworth expansion for $\sigma^{-1}\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\lambda'\Pi X_g'u_g$ involves the average squared variance

$$\theta = \frac{1}{G}\sum_{g=1}^{G}(\sigma_g^2)^2 = \frac{1}{G}\sum_{g=1}^{G}\left(\lambda'\Pi X_g' E[u_g u_g' \mid \mathbf{X}=\mathbf{x}] X_g \Pi\lambda\right)^2$$

and the error variance matrix $E[u_g u_g' \mid \mathbf{X}=\mathbf{x}]$ is not specified in this paper. To illustrate how the average squared variance appears in the Cramér-Edgeworth expansion, we assume $\sigma=1$. In order to approximate the distribution of $\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\lambda'\Pi X_g'u_g$ up to $o(G^{-1})$, we characterize $\log(E[\exp(iz\lambda'\Pi X_g'u_g/\sqrt{G})])$ up to $o(G^{-2})$. We have the Taylor expansion of

$$E[\exp(iz\lambda'\Pi X_g'u_g/\sqrt{G})\mid \mathbf{X}=\mathbf{x}] = 1+\sigma_g^2\frac{(iz/\sqrt{G})^2}{2!}+\cdots.$$

Since $\log(1+z) = z-\frac{1}{2}z^2+O(\|z\|^3)$ as $z\to 0$, we have

$$\log(E[\exp(iz\lambda'\Pi X_g'u_g/\sqrt{G})\mid \mathbf{X}=\mathbf{x}]) = \sigma_g^2\frac{(iz/\sqrt{G})^2}{2!} - \frac{1}{2}\left(\sigma_g^2\frac{(iz/\sqrt{G})^2}{2!}\right)^2+\cdots.$$

By the independence across g, the log characteristic function of $\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\lambda'\Pi X_g'u_g$ is the sum of the log characteristic function of $\lambda'\Pi X_g'u_g/\sqrt{G}$ and therefore it has the average squared variance $\frac{1}{G}\sum_{g=1}^{G}(\sigma_g^2)^2$. In the above equation, the squared variance σ_g^2 appears because the logarithm function is nonlinear and we use the quadratic approximation to achieve the remainder term of $o(G^{-1})$.

[8] $f_{t'\eta_g|\mathbf{X}=\mathbf{x}}$ is the probability density function of $t'\eta_g$ given $\mathbf{X}=\mathbf{x}$ for every g with $Var(t'\eta_g\mid\mathbf{X}=\mathbf{x})>0$. For

The proof is given in Section A.3.

The above three assumptions constitute a sufficient condition for Assumption 2.

Theorem 3. *Assumptions 3-5 imply Assumption 2.*

The proof is given in Section A.4.

3 Monte Carlo simulations

In this section, we investigate the finite-sample performance of the critical value proposed in Section 2 using simulated data. We compare our method (denoted by "Analytical" in the figures) with a few existing methods, such as (i) the t_{G-1} critical value ("Student"), (ii) the restricted wild cluster bootstrap with Rademacher weights by Cameron et al. (2008) ("CWB"), and (iii) the pairs percentile-t cluster bootstrap ("Pairs"). We use 10,000 simulations and 1000 draws for bootstrap procedures. We consider two designs that are challenging for existing methods but can be accommodated well using our analytic corrections. The first design features binary regressors, which is challenging for the pairs cluster bootstrap, while the second design features skewed errors, which is challenging for the cluster wild bootstrap.

3.1 Design with Binary Regressors

In this section, we follow the simulation design from Bertrand et al. (2004) and Cameron et al. (2008, Section 5.A). It uses a state-year panel of excess earnings from 1979 to 1999 based on the Current Population Survey.[9] For each simulation draw, we randomly select G out of 50 states with replacement. We randomly select the policy change time uniformly from $\{1984, \ldots, 1993\}$ and assume that half of the G states experience the policy change after the selected time period. We construct the policy dummy variable accordingly. By definition, this policy dummy variable has a zero coefficient in the population. We regress the excess earnings on the policy dummy, the year dummies, and the state dummies, and conduct the significance test for the coefficient of the policy dummy variable.

In the data generating process of this section, the skewness of the score is close to zero (with $\hat{\mu}_{1,1,1} = 0.02$ for $G = 10^4$). Although the Cramér condition does not hold for this design because all the variables are discrete (cf., Bhattacharya and Rao, 1976, Ch.5), we can consider whether a critical value accounts for the skewness and kurtosis of the t-statistic, which are the key components of the second-order Cramér-Edgeworth expansion. Our proposed method matches these moments by estimating them explicitly. At the same time, the wild cluster bootstrap approximates the

every t, the distribution of $t'\eta_g$ given $\mathbf{X} = \mathbf{x}$ is either a point mass of zero or absolutely continuous (with respect to the 1-dimensional Lebesgue measure). Therefore, if $Var(t'\eta_g \mid \mathbf{X} = \mathbf{x}) > 0$ then the distribution of $t'\eta_g$ given $\mathbf{X} = \mathbf{x}$ has a probability density function.

[9] We use the data from the replication package of Cameron and Miller (2015): `https://cameron.econ.ucdavis.edu/research/papers.html` In particular, this simulation exercise uses the variable `lnwage` from `CPS_panel.dta` from 1979 to 1999.

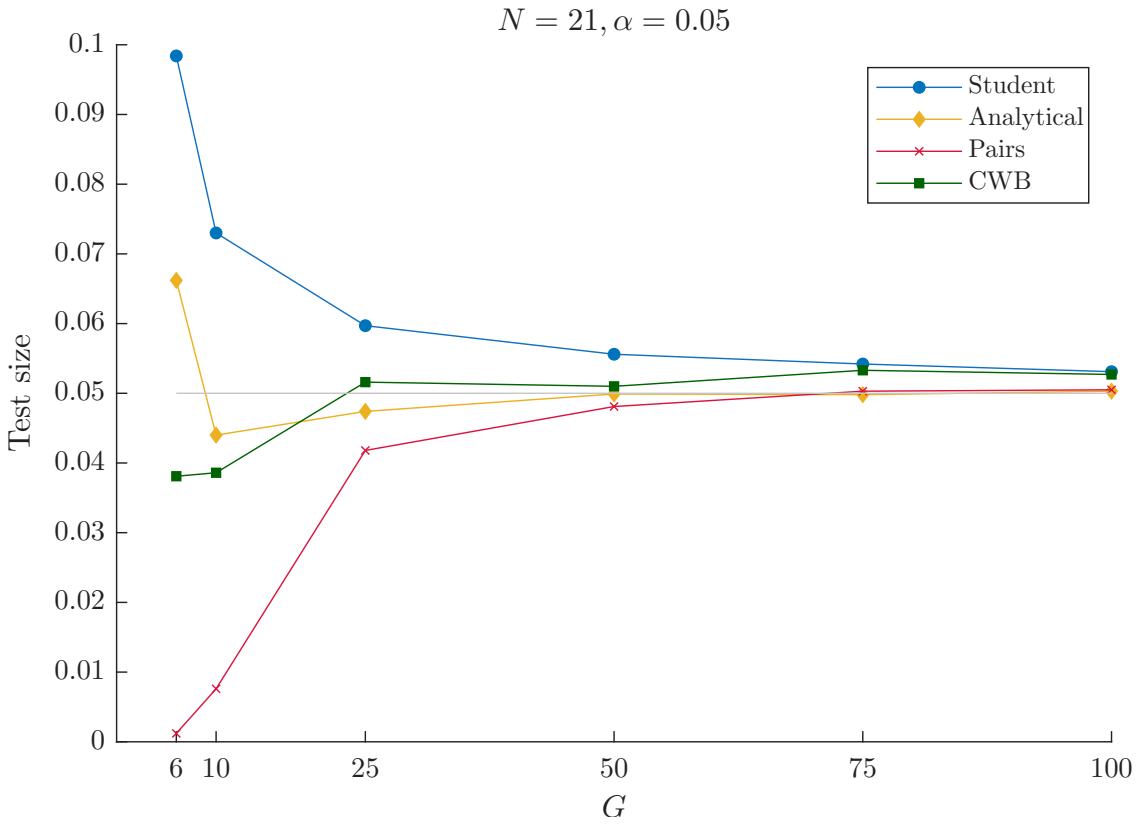


Figure 1: Rejection probabilities for two-sided tests in the Bertrand et al. (2004) design. N is the number of observations per cluster, α is nominal test size.

skewness and kurtosis well because it uses zero skewness and estimates the kurtosis consistently (Djogbenou et al., 2019, Section 5.2).

Figure 1 shows the rejection probabilities for different methods. As documented in Cameron et al. (2008), the pairs cluster bootstrap under-rejects for small values of G (e.g., $G = 6, 10$), while the inference based on the t_{G-1} critical value over-rejects. All methods control the size approximately when G is sufficiently large (e.g., $G = 50$). Our proposed method exhibits comparable performance to the wild cluster bootstrap, even for a small value of $G = 10$.

3.2 Design with a Skewed Error Distribution

To compare the methods when the error has large skewness, we consider the case with $N_g = 1$, $X_{ig} = 1$, and Y_{ig} follows the exponential distribution with unit mean. This design is used in Section 3 of Hall (1983) for one-sided tests. Since the error has a skewness of 2, Theorem 5.2 of Djogbenou et al. (2019) implies that the wild cluster bootstrap does not have asymptotic refinement in this case.

Figure 2 shows the rejection probabilities for different methods with the skewed error distribution. Again, the rejection probabilities of all the methods approach the prespecified significance level (5%) as G increases, which confirms their asymptotic validity. However, the finite-sample perfor-

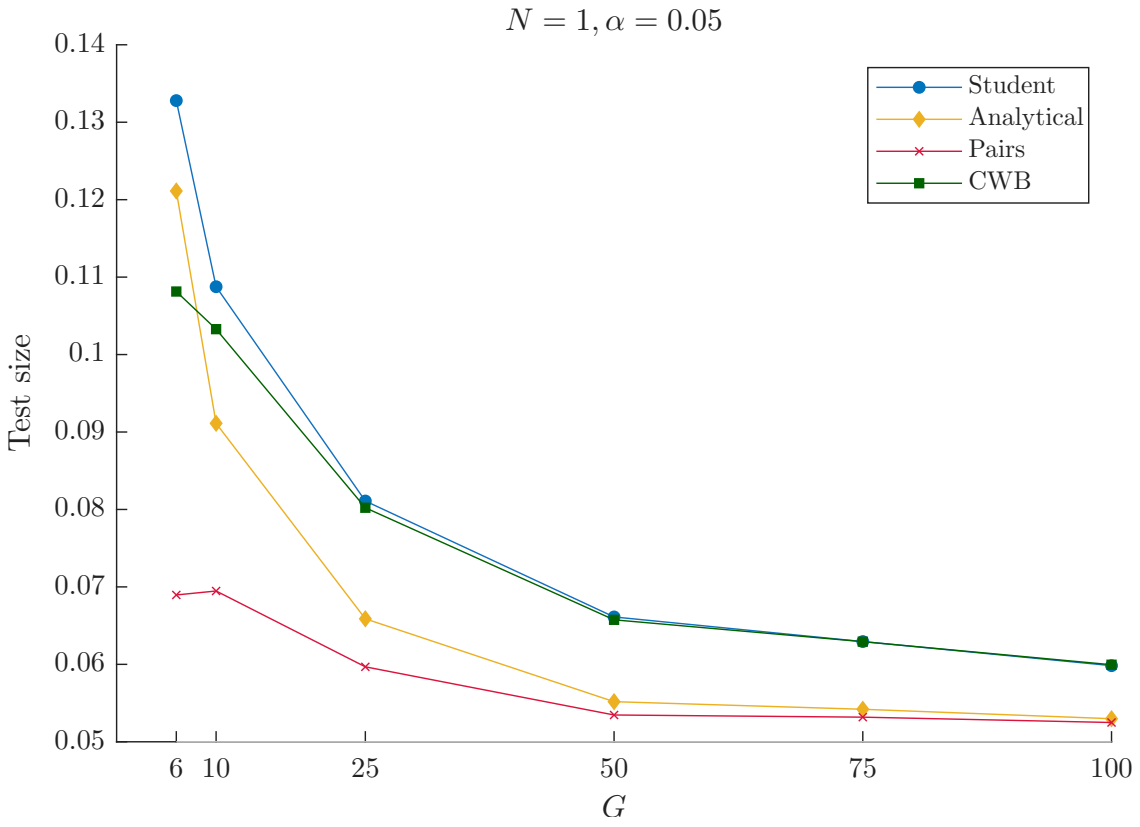


Figure 2: Rejection probabilities for two-sided tests with a skewed error distribution. N is the number of observations per cluster, α is nominal test size.

mance differs. The rejection probabilities of our proposed method approach 5% faster than the wild cluster bootstrap. This is consistent with the theoretical fact that the wild cluster bootstrap does not have asymptotic refinement in this data distribution. In contrast, our analytical approach and the pairs cluster bootstrap both achieve the asymptotic refinement. It explains why these methods are similar to each other and much closer to nominal size than the other two methods in Figure 2.

4 Conclusion

In this paper, we propose an inference method for linear regression with clustered errors, and it achieves third-order asymptotic refinements. Unlike the cluster pairs bootstrap, it does not resample the Gram matrix of $\frac{1}{G}\sum_{g=1}^{G} X_g' X_g$, thus avoiding the small-sample issues of the cluster pairs bootstrap (cf., Cameron et al., 2008). Our simulation results show favorable finite-sample performance of the proposed method. Notably, it works comparably to the wild bootstrap in the simulation design based on Bertrand et al. (2004) and for some designs (with skewed distributions) it has better size control than the wild bootstrap.

References

Angst, J. and G. Poly (2017): "A weak Cramér condition and application to Edgeworth expansions," *Electronic Journal of Probability*, 22, 1–24.

Bai, Z. and C. R. Rao (1991): "Edgeworth expansion of a function of sample means," *The Annals of Statistics*, 1295–1315.

Bertrand, M., E. Duflo, and S. Mullainathan (2004): "How much should we trust differences-in-differences estimates?" *The Quarterly journal of economics*, 119, 249–275.

Bester, C. A., T. G. Conley, and C. B. Hansen (2011): "Inference with dependent data using cluster covariance estimators," *Journal of Econometrics*, 165, 137–151.

Bhattacharya, R. N. and J. K. Ghosh (1978): "On the validity of the formal Edgeworth expansion," *Ann. Statist*, 6, 434–451.

Bhattacharya, R. N. and R. R. Rao (1976): *Normal Approximation and Asymptotic Expansions*, SIAM.

Bobkov, S., G. Chistyakov, and F. Götze (2012): "Bounds for characteristic functions in terms of quantiles and entropy," *Electronic Communications in Probability*, 17, 1–9.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008): "Bootstrap-based improvements for inference with clustered errors," *The review of economics and statistics*, 90, 414–427.

Cameron, A. C. and D. L. Miller (2015): "A practitioner's guide to cluster-robust inference," *Journal of human resources*, 50, 317–372.

——— (2025): "Inference for Regression with Clustered or Spatially Correlated Data," *Working Paper*.

Canay, I. A., A. Santos, and A. M. Shaikh (2021): "The wild bootstrap with a "small" number of "large" clusters," *Review of Economics and Statistics*, 103, 346–363.

Cramér, H. (1928): "On the Composition of Elementary Errors," *Skandinavisk Aktuarietidskrift*, 11, 13–74, 141–180.

Djogbenou, A. A., J. G. MacKinnon, and M. Ø. Nielsen (2019): "Asymptotic Theory and Wild Bootstrap Inference with Clustered Errors," *Journal of Econometrics*, 212, 393–412.

Edgeworth, F. Y. (1905): "The Law of Error," *Proceedings of the Cambridge Philosophical Society*, 20, 36–65.

Hall, P. (1983): "Inverting an Edgeworth expansion," *Annals of Statistics*, 569–576.

——— (2013): *The Bootstrap and Edgeworth Expansion*, Springer Science & Business Media.

Hansen, B. E. (2021): "The exact distribution of the White t-ratio," *Manuscript, University of Wisconsin*.

——— (2025): "Jackknife standard errors for clustered regression," *Review of Economic Studies*.

Horn, R. A. and C. R. Johnson (2012): *Matrix Analysis*, Cambridge University Press.

Imbens, G. W. and M. Kolesar (2016): "Robust standard errors in small samples: Some practical advice," *Review of Economics and Statistics*, 98, 701–712.

Liu, R. Y. (1988): "Bootstrap procedures under some non-iid models," *The annals of statistics*, 1696–1708.

MacKinnon, J. G., M. Ø. Nielsen, and M. D. Webb (2023): "Cluster-robust inference: A guide to empirical practice," *Journal of Econometrics*, 232, 272–299.

McCaffrey, D. F. and R. M. Bell (2003): "Bias reduction in standard errors for linear regression with multi-stage samples," *Quality control and applied statistics*, 48, 677–682.

Student (1908): "The probable error of a mean," *Biometrika*, 1–25.

Young, A. (2016): "Improved, nearly exact, statistical inference with robust and clustered covariance matrices using effective degrees of freedom corrections," *Manuscript, London School of Economics*.

Appendix

A Proofs

A.1 Proof of Lemma 2.1

Proof of Lemma 2.1. Note that

$$\lambda'\Pi X_g'\hat{u}_g = \lambda'\Pi X_g'u_g - \lambda'\Pi X_g'X_g(\hat{\beta}-\beta)$$

and that

$$\begin{aligned}
\frac{1}{G}\sum_{g=1}^{G}(\lambda'\Pi X_g'\hat{u}_g)^2 &= \frac{1}{G}\sum_{g=1}^{G}(\lambda'\Pi X_g'u_g)^2 \\
&\quad - 2\frac{1}{G}\sum_{g=1}^{G}(\lambda'\Pi X_g'u_g)'\lambda'\Pi X_g'X_g(\hat{\beta}-\beta) \\
&\quad + \frac{1}{G}\sum_{g=1}^{G}(\hat{\beta}-\beta)'X_g'X_g\Pi\lambda\lambda'\Pi X_g'X_g(\hat{\beta}-\beta) \\
&= \frac{1}{G}\sum_{g=1}^{G}(\lambda'\Pi X_g'u_g)^2 \\
&\quad - 2\left(\frac{1}{G}\sum_{g=1}^{G}(X_g'X_g\Pi\lambda\lambda'\Pi X_g'u_g)\right)(\hat{\beta}-\beta) \\
&\quad + \frac{1}{G}(\hat{\beta}-\beta)'\left(\sum_{g=1}^{G}X_g'X_g\Pi\lambda\lambda'\Pi X_g'X_g\right)(\hat{\beta}-\beta).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\hat{\sigma}^2/\sigma^2 &= 1 + \frac{1}{G}\sum_{g=1}^{G}\omega_{3g} \\
&\quad - 2\sigma^{-2}\left(\frac{1}{G}\sum_{g=1}^{G}X_g'X_g\Pi\lambda\lambda'\Pi X_g'u_g\right)'\left(\frac{1}{G}\sum_{g=1}^{G}\Pi X_g'u_g\right) \\
&\quad + \sigma^{-2}\left(\frac{1}{G}\sum_{g=1}^{G}\Pi X_g'u_g\right)'\frac{1}{G}\sum_{g=1}^{G}X_g'X_g\Pi\lambda\lambda'\Pi X_g'X_g\left(\frac{1}{G}\sum_{g=1}^{G}\Pi X_g'u_g\right).
\end{aligned}$$

□

A.2 Proof of Theorem 1

Lemma A.1. *Let $\zeta_1, \ldots, \zeta_G$ be any independent random variables with a bounded average fourth central moment. For any sequence $\{\varepsilon_G\}$ with $\varepsilon_G > 0$ and $G\varepsilon_G^4 \to \infty$, we have*

$$Pr\left(\left|\frac{1}{G}\sum_{g=1}^{G}(\zeta_g - E[\zeta_g \mid \mathbf{X}=\mathbf{x}]\right)\right| > \varepsilon_G \mid \mathbf{X}=\mathbf{x}) = o(G^{-1}).$$

Proof. By Markov's inequality (for the fourth moment), we have

$$\begin{aligned}
&Pr\left(\left|\frac{1}{G}\sum_{g=1}^{G}(\zeta_g - E[\zeta_g \mid \mathbf{X}=\mathbf{x}])\right| > \varepsilon_G \mid \mathbf{X}=\mathbf{x}\right) \\
&\leq \frac{E\left[\left(\frac{1}{G}\sum_{g=1}^{G}(\zeta_g - E[\zeta_g \mid \mathbf{X}=\mathbf{x}])\right)^4 \mid \mathbf{X}=\mathbf{x}\right]}{\varepsilon_G^4} \\
&= \frac{\frac{1}{G^4}\sum_{g=1}^{G} E[(\zeta_{g_1} - E[\zeta_{g_1} \mid \mathbf{X}=\mathbf{x}])^4 \mid \mathbf{X}=\mathbf{x}]}{\varepsilon_G^4} \\
&+ 6\frac{\frac{1}{G^4}\sum_{g_1=1}^{G}\sum_{g_2=1}^{G} E[(\zeta_{g_1} - E[\zeta_{g_1} \mid \mathbf{X}=\mathbf{x}])^2(\zeta_{g_2} - E[\zeta_{g_2} \mid \mathbf{X}=\mathbf{x}])^2 \mid \mathbf{X}=\mathbf{x}]}{\varepsilon_G^4},
\end{aligned}$$

where the equality follows from the independence across g. Note that

$$\begin{aligned}
&\frac{1}{G^4}\sum_{g_1=1}^{G}\sum_{g_2=1}^{G} E[(\zeta_{g_1} - E[\zeta_{g_1} \mid \mathbf{X}=\mathbf{x}])^2(\zeta_{g_2} - E[\zeta_{g_2} \mid \mathbf{X}=\mathbf{x}])^2 \mid \mathbf{X}=\mathbf{x}] \\
&\leq \frac{1}{G^4}\sum_{g_1=1}^{G}\sum_{g_2=1}^{G} E[(\zeta_{g_1} - E[\zeta_{g_1} \mid \mathbf{X}=\mathbf{x}])^4 \mid \mathbf{X}=\mathbf{x}]^{1/2} E[(\zeta_{g_2} - E[\zeta_{g_2} \mid \mathbf{X}=\mathbf{x}])^4 \mid \mathbf{X}=\mathbf{x}]^{1/2} \\
&= \frac{1}{G^2}(\frac{1}{G}\sum_{g=1}^{G} E[(\zeta_g - E[\zeta_g \mid \mathbf{X}=\mathbf{x}])^4 \mid \mathbf{X}=\mathbf{x}]^{1/2})^2 \\
&\leq \frac{1}{G^2}\frac{1}{G}\sum_{g=1}^{G} E[(\zeta_g - E[\zeta_g \mid \mathbf{X}=\mathbf{x}])^4 \mid \mathbf{X}=\mathbf{x}],
\end{aligned}$$

where the second inequality follows from the Cauchy-Schwarz inequality, and the last inequality follows from Jensen's inequality. Since $G\varepsilon_G^4 \to \infty$, we have the statement of this lemma. □

Lemma A.2. *Let $\hat{\vartheta}_{1,G}, \ldots, \hat{\vartheta}_{L,G}$ be L sequences of random variables. Suppose $\hat{\rho}_G$ is the composition of a finite number of matrix additions and/or matrix multiplications of $\hat{\vartheta}_{1,G}, \ldots, \hat{\vartheta}_{L,G}$. Suppose that the sequence $\sum_{\ell=1,\ldots,L} \|\vartheta_{\ell,G}\|$ is bounded and $\sum_{\ell=1}^{L} Pr(\|\hat{\vartheta}_{\ell,G} - \vartheta_{\ell,G}\| > \varepsilon_G \mid \mathbf{X}=\mathbf{x}) = o(G^{-1})$ for some diminishing sequence $\{\varepsilon_G\}$. Then there is some constant C such that*

$$Pr(\|\hat{\rho}_G - \rho_G\| > C\varepsilon_G \mid \mathbf{X}=\mathbf{x}) = o(G^{-1}).$$

If, in addition, ρ_G is an invertible matrix whose singular values are bounded away from zero, then there is some constant C such that

$$Pr(\|\hat{\rho}_G^{-1} - \rho_G^{-1}\| > C\varepsilon_G \mid \mathbf{X} = \mathbf{x}) = o(G^{-1}).$$

Proof. For the first result, it suffices to show that $Pr(\|(\hat{\vartheta}_{1,G}\hat{\vartheta}_{2,G} + \hat{\vartheta}_{3,G}) - (\vartheta_{1,G}\vartheta_{2,G} + \vartheta_{3,G})\| > \varepsilon_G \mid \mathbf{X} = \mathbf{x}) = o(G^{-1})$ for some constant C. Since

$$\begin{aligned}\|(\hat{\vartheta}_{1,G}\hat{\vartheta}_{2,G} + \hat{\vartheta}_{3,G}) - (\vartheta_{1,G}\vartheta_{2,G} + \vartheta_{3,G})\| \leq{}& \|\hat{\vartheta}_{1,G} - \vartheta_{1,G}\|\|\hat{\vartheta}_{2,G} - \vartheta_{2,G}\| + \|\vartheta_{1,G}\|\|\hat{\vartheta}_{2,G} - \vartheta_{2,G}\| \\ &+ \|\hat{\vartheta}_{1,G} - \vartheta_{1,G}\|\|\vartheta_{2,G}\| + \|\hat{\vartheta}_{3,G} - \vartheta_{3,G}\|,\end{aligned}$$

we have

$$\begin{aligned}&Pr(\|(\hat{\vartheta}_{1,G}\hat{\vartheta}_{2,G} + \hat{\vartheta}_{3,G}) - (\vartheta_{1,G}\vartheta_{2,G} + \vartheta_{3,G})\| > \varepsilon_G^2 + (\|\vartheta_{1,G}\| + \|\vartheta_{2,G}\|)\varepsilon_G + \varepsilon_G \mid \mathbf{X} = \mathbf{x}) \\ &\leq Pr(\|\hat{\vartheta}_{1,G} - \vartheta_{1,G}\|\|\hat{\vartheta}_{2,G} - \vartheta_{2,G}\| > \varepsilon_G^2 \mid \mathbf{X} = \mathbf{x}) \\ &\quad + Pr(\|\vartheta_{1,G}\|\|\hat{\vartheta}_{2,G} - \vartheta_{2,G}\| > \|\vartheta_{1,G}\|\varepsilon_G \mid \mathbf{X} = \mathbf{x}) + Pr(\|\hat{\vartheta}_{1,G} - \vartheta_{1,G}\|\|\vartheta_{2,G}\| > \|\vartheta_{2,G}\|\varepsilon_G \mid \mathbf{X} = \mathbf{x}) \\ &\quad + Pr(\|\hat{\vartheta}_{3,G} - \vartheta_{3,G}\| > \varepsilon_G \mid \mathbf{X} = \mathbf{x}) \\ &\leq Pr(\|\hat{\vartheta}_{1,G} - \vartheta_{1,G}\| > \varepsilon_G \mid \mathbf{X} = \mathbf{x}) + Pr(\|\hat{\vartheta}_{2,G} - \vartheta_{2,G}\| > \varepsilon_G \mid \mathbf{X} = \mathbf{x}) \\ &\quad + Pr(\|\hat{\vartheta}_{2,G} - \vartheta_{2,G}\| > \varepsilon_G \mid \mathbf{X} = \mathbf{x}) + Pr(\|\hat{\vartheta}_{1,G} - \vartheta_{1,G}\| > \varepsilon_G \mid \mathbf{X} = \mathbf{x}) \\ &\quad + Pr(\|\hat{\vartheta}_{3,G} - \vartheta_{3,G}\| > \varepsilon_G \mid \mathbf{X} = \mathbf{x}) \\ &= o(G^{-1}).\end{aligned}$$

By taking C such that $C > \varepsilon_G + (\|\vartheta_{1,G}\| + \|\vartheta_{2,G}\|) + 1$, the first result holds. Now we will show the second result about the inverse as follows. By Horn and Johnson (2012, Eq.(5.8.6)), we have

$$\|\hat{\rho}_G^{-1} - \rho_G^{-1}\| \leq \|\rho_G^{-1}\| \frac{\|\rho_G^{-1}\|\|\hat{\rho}_G - \rho_G\|}{1 - \|\rho_G^{-1}\|\|\hat{\rho}_G - \rho_G\|}$$

as long as $\|\rho_G^{-1}\|\|\hat{\rho}_G - \rho_G\| < 1$. When $\varepsilon_G \leq 0.5\|\rho_G^{-1}\|^{-1}$, we have

$$\|\hat{\rho}_G - \rho_G\| < \varepsilon_G \implies \|\hat{\rho}_G^{-1} - \rho_G^{-1}\| \leq \|\rho_G^{-1}\| \frac{\|\rho_G^{-1}\|\varepsilon_G}{1 - \|\rho_G^{-1}\|\varepsilon_G} \leq 2\|\rho_G^{-1}\|^2\varepsilon_G.$$

Therefore, the second result holds. □

Lemma A.3. *Suppose the assumptions in Theorem 1 as well as the null hypothesis H_0. For every $z > 0$, there is some sequence $\{\varepsilon_G\}$ such that $\varepsilon_G = o(1)$ and that $Pr(|\hat{q}_2(z) - q_2(z)| > \varepsilon_G \mid \mathbf{X} = \mathbf{x}) = o(G^{-1})$.*

Proof. By Lemma A.2 and Assumption 1, it suffices to investigate $\hat{\mu}_{1,2} - \mu_{1,2}$, $\hat{\mu}_{2,2} - \mu_{2,2}$, $\hat{\mu}_{1,1,1} - \mu_{1,1,1}$,

and $\hat{\mu}_{1,1,1,1} - \mu_{1,1,1,1}$. Note that

$$\begin{aligned}
\hat{\mu}_{1,2} - \mu_{1,2} &= \frac{1}{G}\sum_g ((\hat{\omega}_{1g_1} - \omega_{1g_1})\omega_{2g_1} + \omega_{1g_1}(\hat{\omega}_{2g_1} - \omega_{2g_1}) + (\hat{\omega}_{1g_1} - \omega_{1g_1})(\hat{\omega}_{2g_1} - \omega_{2g_1})) \\
&\quad + \frac{1}{G}\sum_g (\omega_{1g_1}\omega_{2g_1} - E[\omega_{1g_1}\omega_{2g_1} \mid \mathbf{X} = \mathbf{x}]) \\
\hat{\mu}_{2,2} - \mu_{2,2} &= \frac{1}{G}\sum_g ((\hat{\omega}_{2g} - \omega_{2g})'\Gamma\omega_{2g} + \omega_{2g}'\Gamma(\hat{\omega}_{2g} - \omega_{2g}) + (\hat{\omega}_{2g} - \omega_{2g})'\Gamma(\hat{\omega}_{2g} - \omega_{2g})) \\
&\quad + \frac{1}{G}\sum_g (\omega_{2g}'\Gamma\omega_{2g} - E[\omega_{2g}'\Gamma\omega_{2g} \mid \mathbf{X} = \mathbf{x}]), \\
\hat{\mu}_{1,1,1} - \mu_{1,1,1} &= \frac{1}{G}\sum_g (\hat{\omega}_{1g} - \omega_{1g})((\hat{\omega}_{1g} - \omega_{1g})^2 + 3(\hat{\omega}_{1g} - \omega_{1g})\omega_{1g} + 3\omega_{1g}^2) \\
&\quad + \frac{1}{G}\sum_g (\omega_{1g}^3 - E[\omega_{1g}^3 \mid \mathbf{X} = \mathbf{x}]), \\
\hat{\mu}_{1,1,1,1} - \mu_{1,1,1,1} &= \frac{1}{G}\sum_g (\hat{\omega}_{1g} - \omega_{1g})(4\omega_{1g}^3 + 6\omega_{1g}^2(\hat{\omega}_{1g} - \omega_{1g}) + 4\omega_{1g}(\hat{\omega}_{1g} - \omega_{1g})^2 + (\hat{\omega}_{1g} - \omega_{1g})^3) \\
&\quad + \frac{1}{G}\sum_g (\omega_{1g}^4 - E[\omega_{1g}^4 \mid \mathbf{X} = \mathbf{x}]),
\end{aligned}$$

$$\begin{aligned}
\hat{\omega}_{1g} - \omega_{1g} &= ((\hat{\sigma}/\sigma)^{-1} - 1)\omega_{1g} - (\hat{\sigma}/\sigma)^{-1}\sigma^{-1}\lambda'\Pi X_g'X_g(\hat{\beta} - \beta), \\
\hat{\omega}_{2g} - \omega_{2g} &= ((\hat{\sigma}/\sigma)^{-1} - 1)\omega_{2g} - (\hat{\sigma}/\sigma)^{-1}\sigma^{-1}\begin{pmatrix} I_k \\ X_g'X_g\Pi'\lambda\lambda' \end{pmatrix}\Pi X_g'X_g(\hat{\beta} - \beta).
\end{aligned}$$

Each term in the above expressions is bounded as follows:

$$\|\hat{\beta} - \beta\| \leq \|\Pi\| \left\| \frac{1}{G}\sum_{g=1}^{G} X_g'u_g \right\|,$$

$$\begin{aligned}
|\hat{\omega}_{1g} - \omega_{1g}| &\leq |(\hat{\sigma}/\sigma)^{-1} - 1||\omega_{1g}| \\
&\quad + \|\hat{\beta} - \beta\||\hat{\sigma}/\sigma|^{-1}\sigma^{-1}\|\lambda\|\|\Pi\|\|X_g'X_g\|, \\
\|\hat{\omega}_{2g} - \omega_{2g}\| &\leq |(\hat{\sigma}/\sigma)^{-1} - 1|\|\omega_{2g}\| \\
&\quad + \|\hat{\beta} - \beta\||\hat{\sigma}/\sigma|^{-1}\sigma^{-1}\|\Pi\|\|X_g'X_g\| \\
&\quad + \|\hat{\beta} - \beta\||\hat{\sigma}/\sigma|^{-1}\sigma^{-1}\|\lambda\|^2\|\Pi\|^2\|X_g'X_g\|^2,
\end{aligned}$$

$$\begin{aligned}
|\omega_{1g}| &\leq \sigma^{-1}\|\lambda\|\|\Pi\|\|X_g'u_g\|, \\
\|\omega_{2g}\| &\leq \sigma^{-1}\|\Pi\|\|X_g'u_g\| + \sigma^{-1}\|\lambda\|^2\|\Pi\|^2\|X_g'X_g\|\|X_g'u_g\|.
\end{aligned}$$

By Lemma A.1-A.2 and Assumption 3, we have the statement of this lemma. □

Proof of Theorem 1. This proof essentially follows the proof of Hall (1983, Theorem 1), while not using his i.i.d. assumption. Let $z = \Phi^{-1}(1-\alpha/2)$. Using the sequence ε_G in Lemma A.3, we have

$$\begin{aligned}
Pr(|t| \le z - G^{-1}\hat{q}_2(z) \mid \mathbf{X}=\mathbf{x}) &= Pr(|t| \le z - G^{-1}\hat{q}_2(z), |\hat{q}_2(z) - q_2(z)| > \varepsilon_G \mid \mathbf{X}=\mathbf{x}) \\
&\quad + Pr(|t| \le z - G^{-1}\hat{q}_2(z), |\hat{q}_2(z) - q_2(z)| \le \varepsilon_G \mid \mathbf{X}=\mathbf{x}) \\
&\le Pr(|\hat{q}_2(z) - q_2(z)| > \varepsilon_G \mid \mathbf{X}=\mathbf{x}) \\
&\quad + Pr(|t| \le z - G^{-1}q_2(z) + G^{-1}\varepsilon_G \mid \mathbf{X}=\mathbf{x}) \\
&= o(G^{-1}) + Pr(|t| \le z - G^{-1}q_2(z) + G^{-1}\varepsilon_G \mid \mathbf{X}=\mathbf{x}). \qquad (1)
\end{aligned}$$

Similarly, we have

$$Pr(|t| \le z - G^{-1}\hat{q}_2(z) \mid \mathbf{X}=\mathbf{x}) \ge o(G^{-1}) + Pr(|t| \le z - G^{-1}q_2(z) - G^{-1}\varepsilon_G \mid \mathbf{X}=\mathbf{x}). \qquad (2)$$

By Assumption 2, we have

$$\begin{aligned}
Pr(|t| \le z - G^{-1}q_2(z) + G^{-1}\varepsilon_G \mid \mathbf{X}=\mathbf{x}) &= 2\Phi(z - G^{-1}q_2(z) + G^{-1}\varepsilon_G) - 1 \\
&\quad + 2G^{-1}q_2(z - G^{-1}q_2(z) + G^{-1}\varepsilon_G)\phi(z - G^{-1}q_2(z) + G^{-1}\varepsilon_G) \\
&\quad + o(G^{-1}) \\
Pr(|t| \le z - G^{-1}q_2(z) - G^{-1}\varepsilon_G \mid \mathbf{X}=\mathbf{x}) &= 2\Phi(z - G^{-1}q_2(z) - G^{-1}\varepsilon_G) - 1 \\
&\quad + 2G^{-1}q_2(z - G^{-1}q_2(z) - G^{-1}\varepsilon_G)\phi(z - G^{-1}q_2(z) - G^{-1}\varepsilon_G) \\
&\quad + o(G^{-1}).
\end{aligned}$$

Since Φ, ϕ, and q_2 are continuously differentiable and $\varepsilon_G = o(1)$, we have

$$\begin{aligned}
Pr(|t| \le z - G^{-1}q_2(z) + G^{-1}\varepsilon_G \mid \mathbf{X}=\mathbf{x}) &= 2\Phi(z) - 1 + o(G^{-1}) \\
Pr(|t| \le z - G^{-1}q_2(z) - G^{-1}\varepsilon_G \mid \mathbf{X}=\mathbf{x}) &= 2\Phi(z) - 1 + o(G^{-1}).
\end{aligned}$$

Together with Eq.(1)-(2), we have

$$Pr(|t| \le z - G^{-1}\hat{q}_2(z) \mid \mathbf{X}=\mathbf{x}) - (1-\alpha) = o(G^{-1}),$$

which implies the statement of this theorem. □

A.3 Proof of Theorem 2

Proof of Theorem 2. By Bobkov, Chistyakov, and Götze (2012, Theorem 2), there is a positive constant c_3 such that, if $g \in \mathbf{B}(t/\|t\|)$ and $\|t\| > R$, then

$$\begin{aligned}\left|E\left[\exp\left(it'\eta_g\right) \mid \mathbf{X}=\mathbf{x}\right]\right| &= \left|E\left[\exp\left(iVar(t'\eta_g \mid \mathbf{X}=\mathbf{x})^{1/2}(Var(t'\eta_g \mid \mathbf{X}=\mathbf{x})^{-1/2}t'\eta_g)\right) \mid \mathbf{X}=\mathbf{x}\right]\right| \\ &\leq 1-\frac{c_3}{c_1}.\end{aligned}$$

Then, for every t with $\|t\| > R$, we have

$$\frac{1}{G}\sum_{g=1}^{G}\left|E\left[\exp\left(it'\eta_g\right) \mid \mathbf{X}=\mathbf{x}\right]\right| \leq 1-\frac{1}{G}\sum_{g\in\mathbf{B}(t/\|t\|)}\frac{c_3}{c_1} \leq 1-c_2\frac{c_3}{c_1}$$

for sufficiently large G, which implies Assumption 5. □

A.4 Proof of Theorem 3

Lemma A.4. *Define*

$$\begin{aligned}\mu_{1,3} &= \frac{1}{G}\sum_{g} E\left[\omega_{1g_1}\omega_{3g_1} \mid \mathbf{X}=\mathbf{x}\right] \\ \mu_{3,3} &= \frac{1}{G}\sum_{g} E\left[\omega_{3g_1}^2 \mid \mathbf{X}=\mathbf{x}\right], \\ \mu_{1,1,3} &= \frac{1}{G}\sum_{g} E\left[\omega_{1g}^2\omega_{3g} \mid \mathbf{X}=\mathbf{x}\right],\end{aligned}$$

Then

$$\mu_{1,3} = \mu_{1,1,1},\ \mu_{3,3} = \mu_{1,1,1,1} - \theta, \quad \textit{and } \mu_{1,1,3} = \mu_{1,1,1,1} - \theta.$$

Proof. Since

$$\begin{aligned}\omega_{1g} &= \sigma^{-1}\lambda'\Pi X_g' u_g, \\ \omega_{3g} &= \omega_{1g}^2 - E[\omega_{1g}^2 \mid \mathbf{X}=\mathbf{x}],\end{aligned}$$

we have

$$\begin{aligned}
\mu_{1,3} &= \frac{1}{G}\sum_g E\left[\omega_{1g_1}(\omega_{1g}^2 - E[\omega_{1g}^2 \mid \mathbf{X}=\mathbf{x}]) \mid \mathbf{X}=\mathbf{x}\right] \\
&= \frac{1}{G}\sum_g E\left[\omega_{1g_1}^3 \mid \mathbf{X}=\mathbf{x}\right] - \frac{1}{G}\sum_g E\left[\omega_{1g_1} \mid \mathbf{X}=\mathbf{x}\right] E[\omega_{1g}^2 \mid \mathbf{X}=\mathbf{x}] \\
&= \frac{1}{G}\sum_g E\left[\omega_{1g_1}^3 \mid \mathbf{X}=\mathbf{x}\right] \\
&= \mu_{1,1,1} \\
\mu_{1,1,3} &= \frac{1}{G}\sum_g E\left[\omega_{1g}^2(\omega_{1g}^2 - E[\omega_{1g}^2 \mid \mathbf{X}=\mathbf{x}]) \mid \mathbf{X}=\mathbf{x}\right] \\
&= \frac{1}{G}\sum_g E\left[\omega_{1g}^4 \mid \mathbf{X}=\mathbf{x}\right] - \frac{1}{G}\sum_g E[\omega_{1g}^2 \mid \mathbf{X}=\mathbf{x}]^2 \\
&= \mu_{1,1,1,1} - \theta \\
\mu_{3,3} &= \frac{1}{G}\sum_g E\left[(\omega_{1g}^2 - E[\omega_{1g}^2 \mid \mathbf{X}=\mathbf{x}])^2 \mid \mathbf{X}=\mathbf{x}\right], \\
&= \frac{1}{G}\sum_g E\left[\omega_{1g}^4 \mid \mathbf{X}=\mathbf{x}\right] - \frac{1}{G}\sum_g E[\omega_{1g}^2 \mid \mathbf{X}=\mathbf{x}]^2 \\
&= \mu_{1,1,1,1} - \theta.
\end{aligned}$$

□

Lemma A.5. *Define*

$$\tilde{t} = \sqrt{G}W_1\left(1 + \frac{1}{2}W_2'\Gamma W_2 - \frac{1}{2}W_3 + \frac{3}{8}W_3^2\right),$$

where

$$\begin{pmatrix} W_1 \\ W_2 \\ W_3 \end{pmatrix} = \frac{1}{G}\sum_{g=1}^{G}\begin{pmatrix} \omega_{1g} \\ \omega_{2g} \\ \omega_{3g} \end{pmatrix}.$$

Then

$$\begin{aligned}
E[\tilde{t} \mid \mathbf{X}=\mathbf{x}] &= G^{-1/2}\nu_1 + o(G^{-1}) \\
E[\tilde{t}^2 \mid \mathbf{X}=\mathbf{x}] &= 1 + G^{-1}\nu_2 + o(G^{-1}) \\
E[\tilde{t}^3 \mid \mathbf{X}=\mathbf{x}] &= G^{-1/2}\nu_3 + o(G^{-1}) \\
E[\tilde{t}^4 \mid \mathbf{X}=\mathbf{x}] &= 3 + G^{-1}\nu_4 + o(G^{-1}), \\
E[\tilde{t}^j \mid \mathbf{X}=\mathbf{x}] &= o(G^{-1}) \text{ for every } j \geq 5.
\end{aligned}$$

Proof. First, we are going to show the first moment. Note that

$$G^{1/2}E\left[W_1 \mid \mathbf{X} = \mathbf{x}\right] = 0$$

and that

$$\begin{aligned} G^{1/2}E\left[W_1W_3 \mid \mathbf{X} = \mathbf{x}\right] &= G^{-3/2}\sum_{g_1,g_2} E\left[\omega_{1g_1}\omega_{3g_2} \mid \mathbf{X} = \mathbf{x}\right] \\ &= G^{-1/2}\frac{1}{G}\sum_{g} E\left[\omega_{1g}\omega_{3g} \mid \mathbf{X} = \mathbf{x}\right] \\ &= G^{-1/2}\mu_{1,3}, \end{aligned}$$

where the second equality holds because ω_{1g} and ω_{3g} are mean zero. By Djogbenou et al. (2019, Lemma A.3),

$$G^{1/2}E\left[W_1W_{j_1}'W_{j_2} \mid \mathbf{X} = \mathbf{x}\right] = o(G^{-1}).$$

Since

$$\tilde{t} = \sqrt{G}W_1\left(1 + \frac{1}{2}W_2'\Gamma W_2 - \frac{1}{2}W_3 + \frac{3}{8}W_3^2\right),$$

we have

$$E[\tilde{t} \mid \mathbf{X} = \mathbf{x}] = -G^{-1/2}\frac{\mu_{1,3}}{2} + o(G^{-1}).$$

By Lemma A.4, we have the first equation of this lemma.

Second, we are going to show the second moment. Note that

$$GE\left[W_1^2 \mid \mathbf{X} = \mathbf{x}\right] = 1,$$

that

$$\begin{aligned} GE\left[W_1^2W_3 \mid \mathbf{X} = \mathbf{x}\right] &= G^{-2}\sum_{g_1,g_2,g_3} E\left[\omega_{1g_1}\omega_{1g_2}\omega_{3g_3} \mid \mathbf{X} = \mathbf{x}\right] \\ &= G^{-1}\frac{1}{G}\sum_{g} E\left[\omega_{1g}^2\omega_{3g} \mid \mathbf{X} = \mathbf{x}\right] \\ &= G^{-1}\mu_{1,1,3}, \end{aligned}$$

and that

$$
\begin{aligned}
&GE\left[W_1^2 W_{j_1}' W_{j_2} \mid \mathbf{X}=\mathbf{x}\right] \\
&= G^{-3} \sum_{g_1,g_2,g_3,g_4} E\left[\omega_{1g_1}\omega_{1g_2}\omega_{j_1g_3}'\omega_{j_2g_4} \mid \mathbf{X}=\mathbf{x}\right] \\
&= G^{-3} \sum_{g_1} E\left[\omega_{1g_1}^2\omega_{j_1g_1}'\omega_{j_2g_1} \mid \mathbf{X}=\mathbf{x}\right] \\
&+ G^{-3} \sum_{g_1\neq g_2} \left(E\left[\omega_{1g_2}^2 \mid \mathbf{X}=\mathbf{x}\right] E\left[\omega_{j_1g_1}'\omega_{j_2g_1} \mid \mathbf{X}=\mathbf{x}\right] + 2E\left[\omega_{1g_1}\omega_{j_1g_1}' \mid \mathbf{X}=\mathbf{x}\right] E\left[\omega_{1g_2}\omega_{j_2g_2} \mid \mathbf{X}=\mathbf{x}\right]\right) \\
&= G^{-1}\left(\frac{1}{G}\sum_{g_1} E\left[\omega_{j_1g_1}'\omega_{j_2g_1} \mid \mathbf{X}=\mathbf{x}\right] + 2\frac{1}{G}\sum_{g_1} E\left[\omega_{1g_1}\omega_{j_1g_1}' \mid \mathbf{X}=\mathbf{x}\right] \frac{1}{G}\sum_{g_2} E\left[\omega_{1g_2}\omega_{j_2g_2} \mid \mathbf{X}=\mathbf{x}\right]\right) + o(G^{-1}) \\
&= G^{-1}\left(\mu_{j_1,j_2} + 2\mu_{1,j_1}'\mu_{1,j_2}\right) + o(G^{-1}).
\end{aligned}
$$

By Djogbenou et al. (2019, Lemma A.3),

$$
GE\left[W_1^2 W_{j_1}\cdots W_{j_k} \mid \mathbf{X}=\mathbf{x}\right] = o(G^{-1}) \text{ for } k \geq 3.
$$

Since

$$
\begin{aligned}
\tilde{t}^2 &= GW_1^2\left(1 + \frac{1}{2}W_2'\Gamma W_2 - \frac{1}{2}W_3 + \frac{3}{8}W_3^2\right)^2 \\
&= GW_1^2\left(1 - W_3 + W_3^2 + W_2'\Gamma W_2 + (\text{products of three or more terms of } W_2, W_3)\right),
\end{aligned}
$$

we have

$$
E[\tilde{t}^2 \mid \mathbf{X}=\mathbf{x}] = 1 + G^{-1}\left(-\mu_{1,1,3} + (\mu_{3,3} + 2\mu_{1,3}^2) + (\mu_{2,2} + 2\mu_{1,2}'\mu_{1,2})\right) + o(G^{-1}).
$$

By Lemma A.4, we have the second equation of this lemma.

Third, we are going to show the third moment. Note that

$$
\begin{aligned}
G^{3/2}E\left[W_1^3 \mid \mathbf{X}=\mathbf{x}\right] &= G^{-3/2}\sum_{g_1,g_2,g_3} E\left[\omega_{1g_1}\omega_{1g_2}\omega_{1g_3} \mid \mathbf{X}=\mathbf{x}\right] \\
&= G^{-3/2}\sum_{g_1} E\left[\omega_{1g_1}^3 \mid \mathbf{X}=\mathbf{x}\right] \\
&= G^{-1/2}\mu_{1,1,1},
\end{aligned}
$$

and that

$$
\begin{aligned}
&G^{3/2}E\left[W_1^3W_3 \mid \mathbf{X}=\mathbf{x}\right] \\
&= G^{-5/2}\sum_{g_1,g_2,g_3,g_4} E\left[\omega_{1g_1}\omega_{1g_2}\omega_{1g_3}\omega_{3g_4} \mid \mathbf{X}=\mathbf{x}\right] \\
&= G^{-5/2}\sum_{g_1} E\left[\omega_{1g_1}^3\omega_{3g_1} \mid \mathbf{X}=\mathbf{x}\right] \\
&+ G^{-5/2}\sum_{g_1\neq g_2} \left(E\left[\omega_{1g_2}^2 \mid \mathbf{X}=\mathbf{x}\right] E\left[\omega_{1g_1}\omega_{3g_1} \mid \mathbf{X}=\mathbf{x}\right] + 2E\left[\omega_{1g_1}^2 \mid \mathbf{X}=\mathbf{x}\right] E\left[\omega_{1g_2}\omega_{3g_2} \mid \mathbf{X}=\mathbf{x}\right]\right) \\
&= 3G^{-1/2}\frac{1}{G}\sum_{g_1} E\left[\omega_{1g_1}\omega_{3g_1} \mid \mathbf{X}=\mathbf{x}\right] + o(G^{-1}) \\
&= 3G^{-1/2}\mu_{1,3} + o(G^{-1}).
\end{aligned}
$$

By Djogbenou et al. (2019, Lemma A.3),

$$
G^{3/2}E\left[W_1^3W_{j_1}\cdots W_{j_k} \mid \mathbf{X}=\mathbf{x}\right] = o(G^{-1}) \text{ for } k \geq 2.
$$

Since

$$
\begin{aligned}
\tilde{t}^3 &= G^{3/2}W_1^3\left(1+\frac{1}{2}W_2'\Gamma W_2 - \frac{1}{2}W_3 + \frac{3}{8}W_3^2\right)^3 \\
&= G^{3/2}W_1^3\left(1-\frac{3}{2}W_3 + (\text{products of two or more terms of } W_2, W_3)\right),
\end{aligned}
$$

we have

$$
E[\tilde{t}^3 \mid \mathbf{X}=\mathbf{x}] = G^{-1/2}(\mu_{1,1,1} - \frac{9}{2}\mu_{1,3}) + o(G^{-1}).
$$

By Lemma A.4, we have the third equation of this lemma.

Fourth, we are going to show the fourth moment. Note that

$$
\begin{aligned}
G^2E\left[W_1^4 \mid \mathbf{X}=\mathbf{x}\right] &= G^{-2}\sum_{g_1,g_2,g_3,g_4} E\left[\omega_{1g_1}\omega_{1g_2}\omega_{1g_3}\omega_{1g_4} \mid \mathbf{X}=\mathbf{x}\right] \\
&= G^{-2}\sum_{g_1} E\left[\omega_{1g_1}^4 \mid \mathbf{X}=\mathbf{x}\right] + 3\left(\frac{1}{G}\sum_{g_1} E\left[\omega_{1g_1}^2 \mid \mathbf{X}=\mathbf{x}\right]\right)\left(G^{-1}\sum_{g_2\neq g_1} E\left[\omega_{1g_2}^2 \mid \mathbf{X}=\mathbf{x}\right]\right) \\
&= 3 + G^{-1}\frac{1}{G}\sum_{g_1}\left(E\left[\omega_{1g_1}^4 \mid \mathbf{X}=\mathbf{x}\right] - 3E\left[\omega_{1g_1}^2 \mid \mathbf{X}=\mathbf{x}\right]^2\right) \\
&= 3 + G^{-1}(\mu_{1,1,1,1} - 3\theta),
\end{aligned}
$$

that

$$
\begin{aligned}
&G^2 E\left[W_1^4 W_3 \mid \mathbf{X}=\mathbf{x}\right] \\
&= G^{-3} \sum_{g_1,g_2,g_3,g_4,g_5} E\left[\omega_{3g_1}\omega_{1g_2}\omega_{1g_3}\omega_{1g_4}\omega_{1g_5} \mid \mathbf{X}=\mathbf{x}\right] \\
&= G^{-3} \sum_{g_1,g_2} \left(6E\left[\omega_{3g_1}\omega_{1g_1}^2 \mid \mathbf{X}=\mathbf{x}\right] E\left[\omega_{1g_1}^2 \mid \mathbf{X}=\mathbf{x}\right] + 4E\left[\omega_{3g_1}\omega_{1g_1} \mid \mathbf{X}=\mathbf{x}\right] E\left[\omega_{1g_1}^3 \mid \mathbf{X}=\mathbf{x}\right]\right) \\
&+ G^{-3} \sum_{g_1} E\left[\omega_{3g_1}\omega_{1g_1}\omega_{1g_1}\omega_{1g_1}\omega_{1g_1} \mid \mathbf{X}=\mathbf{x}\right] + o(G^{-1}) \\
&= G^{-1}\left(6\mu_{1,1,3} + 4\mu_{1,3}\mu_{1,1,1}\right) + o(G^{-1}),
\end{aligned}
$$

and that

$$
\begin{aligned}
&G^2 E\left[W_1^4 W'_{j_1} W_{j_2} \mid \mathbf{X}=\mathbf{x}\right] \\
&= G^{-4} \sum_{g_1,g_2,g_3,g_4,g_5,g_6} E\left[\omega'_{j_1g_1}\omega_{j_2g_2}\omega_{1g_3}\omega_{1g_4}\omega_{1g_5}\omega_{1g_6} \mid \mathbf{X}=\mathbf{x}\right] \\
&= G^{-4} \sum_{g_1,g_2,g_3} \left(3E\left[\omega'_{j_1g_1}\omega_{j_2g_1}\omega_{1g_2}^2\omega_{1g_3}^2 \mid \mathbf{X}=\mathbf{x}\right] + 12E\left[\omega'_{j_1g_1}\omega_{j_2g_2}\omega_{1g_1}\omega_{1g_2}\omega_{1g_3}^2 \mid \mathbf{X}=\mathbf{x}\right]\right) \\
&+ G^{-4} \sum_{g_1,g_2} \left(4E\left[\omega'_{j_1g_1}\omega_{j_2g_1}\omega_{1g_1}\omega_{1g_2}^3 \mid \mathbf{X}=\mathbf{x}\right] + 6E\left[\omega'_{j_1g_1}\omega_{j_2g_2}\omega_{1g_1}^2\omega_{1g_2}^2 \mid \mathbf{X}=\mathbf{x}\right]\right) \\
&+ G^{-4} \sum_{g_1} E\left[\omega'_{j_1g_1}\omega_{j_2g_1}\omega_{1g_1}\omega_{1g_1}\omega_{1g_1}\omega_{1g_1} \mid \mathbf{X}=\mathbf{x}\right] + o(G^{-1}) \\
&= G^{-4} \sum_{g_1,g_2,g_3} \left(3E\left[\omega'_{j_1g_1}\omega_{j_2g_1}\omega_{1g_2}^2\omega_{1g_3}^2 \mid \mathbf{X}=\mathbf{x}\right] + 12E\left[\omega'_{j_1g_1}\omega_{j_2g_2}\omega_{1g_1}\omega_{1g_2}\omega_{1g_3}^2 \mid \mathbf{X}=\mathbf{x}\right]\right) + o(G^{-1}) \\
&= G^{-1}\left(3\mu_{j_1,j_2} + 12\mu'_{1,j_1}\mu_{1,j_2}\right) + o(G^{-1}).
\end{aligned}
$$

By Djogbenou et al. (2019, Lemma A.3),

$$
G^2 E\left[W_1^4 W_{j_1}\cdots W_{j_k} \mid \mathbf{X}=\mathbf{x}\right] = o(G^{-1}) \text{ for } k \geq 3.
$$

Since

$$
\begin{aligned}
\tilde{t}^4 &= G^2 W_1^4 \left(1 + \frac{1}{2}W_2'\Gamma W_2 - \frac{1}{2}W_3 + \frac{3}{8}W_3^2\right)^4 \\
&= G^2 W_1^4 \left(1 - 2W_3 + 2W_2'\Gamma W_2 + 3W_3^2 + (\text{products of three or more terms of } W_2, W_3)\right),
\end{aligned}
$$

we have

$$
\begin{aligned}
&E[\tilde{t}^4 \mid \mathbf{X}=\mathbf{x}] - 3 \\
&= G^{-1}((\mu_{1,1,1,1} - 3\theta) - 2(6\mu_{1,1,3} + 4\mu_{1,3}\mu_{1,1,1}) + 2\left(3\mu_{2,2} + 12\mu'_{1,2}\mu_{1,2}\right) + 3\left(3\mu_{3,3} + 12\mu'_{1,3}\mu_{1,3}\right)) + o(G^{-1}).
\end{aligned}
$$

By Lemma A.4, we have the fourth equation of this lemma.

Last, the statement with $j \geq 5$ follows from Djogbenou et al. (2019, Lemma A.3). □

Lemma A.6.

$$Pr(|t - \tilde{t}| > (\log G)^{-1} G^{-1} \mid \mathbf{X} = \mathbf{x}) = o(G^{-1}) \text{ under } H_0.$$

Proof. Throughout this proof, we impose H_0. We have

$$t - \tilde{t} = -\sqrt{G} W_1 \left(1 + \frac{1}{2} W_2' \Gamma W_2 - \frac{1}{2} W_3 + \frac{3}{8} W_3^2 - (1 - W_2' \Gamma W_2 + W_3)^{-1/2}\right).$$

By the second-order Taylor expansion of $u \mapsto u^{-1/2}$, we have

$$|(1 - W_2' \Gamma W_2 + W_3)^{-1/2} - (1 - \frac{1}{2}(-W_2' \Gamma W_2 + W_3) + \frac{3}{8}(-W_2' \Gamma W_2 + W_3)^2)| \leq | - W_2' \Gamma W_2 + W_3|^3$$

when $| - W_2' \Gamma W_2 + W_3| < 1 - (\frac{5}{16})^{2/7} \approx 0.72$. Therefore,

$$|t - \tilde{t}| \leq \sqrt{G} |W_1| \left(\frac{3}{8} |(W_2' \Gamma W_2)^2 - 2(W_2' \Gamma W_2) W_3| + | - W_2' \Gamma W_2 + W_3|^3\right)$$

when $| - W_2' \Gamma W_2 + W_3| < 1 - (\frac{5}{16})^{2/7}$. By Markov's inequality and Djogbenou et al. (2019, Lemma A.3), we have

$$Pr(|-W_2' \Gamma W_2 + W_3| > 1 - (\frac{5}{16})^{2/7} \mid \mathbf{X} = \mathbf{x}) \leq (1 - (\frac{5}{16})^{2/7}))^{-4} E[|W_2' \Gamma W_2 + W_3|^4 \mid \mathbf{X} = \mathbf{x}] = O(G^{-2})$$

$$Pr(\sqrt{G}|W_1|(W_2' \Gamma W_2)^2 > (\log G)^{-1} G^{-1} \mid \mathbf{X} = \mathbf{x}) \leq ((\log G)^{-1} G^{-1})^{-2} E[G W_1^2 (W_2' \Gamma W_2)^4 \mid \mathbf{X} = \mathbf{x}] = o(G^{-1})$$

$$Pr(\sqrt{G}|W_1|W_2' \Gamma W_2|W_3| > (\log G)^{-1} G^{-1} \mid \mathbf{X} = \mathbf{x}) \leq ((\log G)^{-1} G^{-1})^{-4} E[G^2 W_1^4 (W_2' \Gamma W_2)^4 W_3^4 \mid \mathbf{X} = \mathbf{x}] = o(G^{-1})$$

$$Pr(\sqrt{G}|W_1|(W_2' \Gamma W_2)^3 > (\log G)^{-1} G^{-1} \mid \mathbf{X} = \mathbf{x}) \leq ((\log G)^{-1} G^{-1})^{-2} E[G W_1^2 (W_2' \Gamma W_2)^6 \mid \mathbf{X} = \mathbf{x}] = o(G^{-1})$$

$$Pr(\sqrt{G}|W_1|(W_2' \Gamma W_2)^2 |W_3| > (\log G)^{-1} G^{-1} \mid \mathbf{X} = \mathbf{x}) \leq ((\log G)^{-1} G^{-1})^{-2} E[G W_1^2 (W_2' \Gamma W_2)^4 W_3^2 \mid \mathbf{X} = \mathbf{x}] = o(G^{-1})$$

$$Pr(\sqrt{G}|W_1|(W_2' \Gamma W_2) W_3^2 > (\log G)^{-1} G^{-1} \mid \mathbf{X} = \mathbf{x}) \leq ((\log G)^{-1} G^{-1})^{-2} E[G W_1^2 (W_2' \Gamma W_2)^2 W_3^4 \mid \mathbf{X} = \mathbf{x}] = o(G^{-1})$$

$$Pr(\sqrt{G}|W_1||W_3|^3 > (\log G)^{-1} G^{-1} \mid \mathbf{X} = \mathbf{x}) \leq ((\log G)^{-1} G^{-1})^{-4} E[G^2 W_1^4 W_3^{12} \mid \mathbf{X} = \mathbf{x}] = o(G^{-1}).$$

They imply the statement of this lemma. □

Lemma A.7. *Define*

$$\mathbb{F}(z) = \Phi(z) + G^{-1/2} q_1(z) \phi(z) + G^{-1} q_2(z) \phi(z)$$

where

$$q_1(z) = -(k_1 + \frac{1}{6} k_3 He_2(z)).$$

Then $\int z^j d\mathbb{F}(z) = E[\tilde{t}^j \mid \mathbf{X} = \mathbf{x}] + o(G^{-1})$ *for every* $j \geq 1$.

Proof. By Lemma A.5 and the definitions of q_1 and q_2, the j-th cumulant of $\mathbb{F}$ is different only in the magnitude of $o(G^{-1})$ from the one of $\tilde{t}$ for $j = 1, 2, 3, 4$ and $o(G^{-1})$ for $j \geq 5$. By Hall (2013, Theorem 2.1), the j-th cumulant of $\tilde{t}$ is $o(G^{-1})$ for $j \geq 5$. Therefore, the cumulants of $\mathbb{F}$ are different only in the magnitude of $o(G^{-1})$ from those of $\tilde{t}$. □

Lemma A.8. $\sup_{z\in\mathbb{R}} |Pr(\tilde{t} \leq z \mid \mathbf{X} = \mathbf{x}) - \mathbb{F}(z)| = o(G^{-1})$.

Proof. The proof structure is similar to Bhattacharya and Ghosh (1978, Theorem 2), but we extend it to the triangular array $\{\eta_g\}$ with potentially non-identical distributions. There is a function H such that

$$\tilde{t} = H\left(\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\eta_g; G^{-1/2}\right).$$

By Angst and Poly (2017, Theorem 4.3),

$$\sup_{z\in\mathbb{R}}\left|Pr(\tilde{t} \leq z \mid \mathbf{X} = \mathbf{x}) - \int_{\{H\left(\mathbf{V}_G^{1/2}u; G^{-1/2}\right)\leq z\}}\left(\sum_{r=0}^{2} G^{-r/2}(\tilde{P}_r(-D)\phi_I)\right)(u)du\right| = o(G^{-1}), \quad (3)$$

where ϕ_I is the multidimensional standard normal pdf and $\tilde{P}_r$ is the polynomial defined in Bhattacharya and Rao (1976, Section 7). By Bhattacharya and Ghosh (1978, Lemma 2.1), there are polynomials $\tilde{q}_1(z)$ and $\tilde{q}_2(z)$ such that

$$\sup_{z\in\mathbb{R}}\left|\int_{\{H\left(\mathbf{V}_G^{1/2}u; G^{-1/2}\right)\leq z\}}\left(\sum_{r=0}^{2} G^{-r/2}(\tilde{P}_r(-D)\phi_I)\right)(u)du - \tilde{\mathbb{F}}(z)\right| = o(G^{-1}) \quad (4)$$

and

$$\int H\left(\mathbf{V}_G^{1/2}u; G^{-1/2}\right)^j\left(\sum_{r=0}^{2} G^{-r/2}(\tilde{P}_r(-D)\phi_I)\right)(u)du = \int z^j d\tilde{\mathbb{F}}(z) + o(G^{-1}) \quad (5)$$

for every integer $j \geq 0$, where $\tilde{\mathbb{F}}(z) = \Phi(z) + G^{-1/2}\tilde{q}_1(z)\phi(z) + G^{-1}\tilde{q}_2(z)\phi(z)$.[10] Since $H\left(\mathbf{V}_G^{1/2}u; G^{-1/2}\right)^j$ is a polynomial of u, its moments are written as the derivatives of the characteristic function at the origin and therefore Bhattacharya and Ghosh (1978, Theorem 9.9) implies

$$E[\tilde{t}^j \mid \mathbf{X} = \mathbf{x}] = \int H\left(\mathbf{V}_G^{1/2}u; G^{-1/2}\right)^j\left(\sum_{r=0}^{2} G^{-r/2}(\tilde{P}_r(-D)\phi_I)\right)(u)du + o(G^{-1}).$$

By Lemma A.7 and Eq. (5), we have

$$\int z^j d\mathbb{F}(z) = E[\tilde{t}^j \mid \mathbf{X} = \mathbf{x}] + o(G^{-1}) = \int z^j d\tilde{\mathbb{F}}(z) + o(G^{-1}),$$

This implies that all the coefficients of $G^{-1/2}q_1(z) + G^{-1}q_2(z)$ and $G^{-1/2}\tilde{q}_1(z) + G^{-1}\tilde{q}_2(z)$ are different

[10] Although Bhattacharya and Ghosh (1978) assumes the identical distributions, their Lemma 2.1 is applicable to the triangular array with non-identical distributions.

only in the magnitude of $o(G^{-1})$. Therefore, $(G^{-1/2}q_1(z) + G^{-1}q_2(z))\phi(z)$ and $(G^{-1/2}\tilde{q}_1(z) + G^{-1}\tilde{q}_2(z))\phi(z)$ are different only in the magnitude of $o(G^{-1})$ uniformly in z, and we have

$$\sup_{z\in\mathbb{R}}\left|\mathbb{F}(z) - \tilde{\mathbb{F}}(z)\right| = o(G^{-1}).$$

Together with Eq. (3) and (4), we have the statement of this lemma. □

Proof of Theorem 3. Under H_0, we have

$$\begin{aligned}
\sup_{z\in\mathbb{R}}|P\left(t \le z \mid \mathbf{X} = \mathbf{x}\right) - \mathbb{F}(z)| &= \sup_{z\in\mathbb{R}}|Pr(t \le z \mid \mathbf{X} = \mathbf{x}) - P(\tilde{t} \le z \mid \mathbf{X} = \mathbf{x})| + o(G^{-1}) \\
&= \sup_{z\in\mathbb{R}}\left(Pr(t \le z < \tilde{t} \mid \mathbf{X} = \mathbf{x}) + Pr(\tilde{t} \le z < t \mid \mathbf{X} = \mathbf{x})\right) + o(G^{-1}) \\
&= \sup_{z\in\mathbb{R}} Pr(z - (\log G)^{-1}G^{-1} < \tilde{t} < z + (\log G)^{-1}G^{-1} \mid \mathbf{X} = \mathbf{x}) + o(G^{-1}) \\
&= \sup_{z\in\mathbb{R}}(\mathbb{F}(z + (\log G)^{-1}G^{-1}) - \mathbb{F}(z - (\log G)^{-1}G^{-1})) + o(G^{-1}) \\
&= o(G^{-1}),
\end{aligned}$$

where the first equality follows from Lemma A.8, the third equality follows from Lemma A.6, and the fourth equality follows from Lemma A.8. □