Online Supplement for *Quantile-based Test for Heterogeneous Treatment Effects*

EunYi Chung[†]
University of Illinois, Urbana–Champaign
eunyi@illinois.edu

Mauricio Olivares
LMU Munich
m.olivares@lmu.de

We organize this online appendix as follows. Section I includes the proofs of the main theorems in the main text, *i.e.*, Theorems 1 and 2. These theorems establish the asymptotic behavior of the permutation test based on the 2SKSQ and the Khmaladze transformed 2SKSQ, respectively. Section II contains the proofs of Lemmas 1 and 2. These lemmas establish the asymptotic null behavior of the 2SKSQ and the Khmaladze transformed 2SKSQ, respectively. In Section III, we provide numerical evidence in showing that a test based on quantiles exhibits better size control in finite samples than one based on CDF comparisons. In Section IV, we describe Bitler, Gelbach, and Hoynes's (2017, BGH) simulated outcomes approach in more detail. We emphasize how their approach is not immune to the Durbin problem and what is the source of the problem. Even though their heuristic approach to the Durbin problem yields a correct conclusion (inadequacy of the CSTE model), we argue the theoretical reasoning behind the simulated outcomes approach does not formally address the problem and, therefore, we cannot claim the asymptotic validity of their permutation test. Section V contains the results of BGH's empirical exercise. We include them here verbatim to highlight how their results are qualitatively the same as ours, despite the presence of an estimated nuisance parameter. Lastly, we conclude this appendix with a brief explanation of how to calculate the martingale-transformed statistic that we use in the construction of our proposed permutation test in Section VI.

[†]The second author acknowledges support from the European Research Council (Starting Grant No. 852332). We are grateful to MDRC for granting access to the experimental data we use in this paper. All errors are our own.

I Proofs of Theorems

Throughout we adopt the following notation, not necessarily introduced in the main text. If ξ is a random variable defined on a probability space $(\Omega, \mathscr{B}, P)$, it is assumed that $\xi_1, \ldots, \xi_N$ are coordinate projections on the product space $(\Omega^N, \mathscr{B}^N, P^N)$, and the expectations are computed for P^N. If auxiliary variables—independent of the ξs—are involved, we use a similar convention. In that case, the underlying probability space is assumed to be of the form $(\Omega^N, \mathscr{B}^N, P^N) \times (\mathcal{Z}, \mathscr{C}, Q)$, with $\xi_1, \ldots, \xi_N$ equal to the coordinate projections on the first N coordinates and the additional variables depending only on the $N + 1$st coordinate.

We view the empirical processes here as random maps into $\ell^\infty(\mathscr{T})$—the space of all bounded functions equipped with the uniform norm—and weak convergence is understood as convergence in distribution in $\ell^\infty(\mathscr{T})$. We assume that the class $\mathscr{T}$ is pointwise measurable (Van der Vaart and Wellner, 1996, Example 2.3.4), ruling out measurability problems with regards to suprema.

Independent of the Zs, let $(\pi(1), \ldots, \pi(N))$ and $(\pi'(1), \ldots, \pi'(N))$ be two independent random permutations of $\{1, \ldots, N\}$. Denote $\boldsymbol{Z}_\pi = (Z_{\pi(1)}, \ldots, Z_{\pi(N)})$; $\boldsymbol{Z}_{\pi'}$ is defined the same way with π replaced by π'

I.1 Proof of Theorem 1

We seek to show the asymptotic behavior of $R^K_{\text{N}}(\cdot)$, the permutation distribution based on the 2SKSQ. Since $\hat{\upsilon}_{\text{N}}(\tau; \boldsymbol{Z})$ is a continuous mapping by the arguments in the proof of Lemma 2, it suffices by the continuous mapping theorem (CMT) for randomization distributions (Chung and Romano, 2016, Lemma A.6) to establishes the asymptotic behavior of the permutation distribution based on $\hat{\upsilon}_{\text{N}}(\tau; \boldsymbol{Z})$.

We begin by recentering as in Remark 2 in the main text. Independent of the $\tilde{Z}$s, let $(\pi(1), \ldots, \pi(N))$ and $(\pi'(1), \ldots, \pi'(N))$ be two independent random permutations of $\{1, \ldots, N\}$. Denote $\tilde{\boldsymbol{Z}}_\pi = (\tilde{Z}_{\pi(1)}, \ldots, \tilde{Z}_{\pi(N)})$; $\tilde{\boldsymbol{Z}}_{\pi'}$ is defined the same way with π replaced by π'. As we argued before, the two-sample quantile process in (6) and (8) in the main text are numerically identical. Therefore, we will establish the asymptotic behavior of the permutation distribution based on $\hat{\upsilon}_{\text{N}}(\tau, \tilde{\boldsymbol{Z}})$.

By Lehmann and Romano (2022, Theorem 17.2.3), it suffices to that

$$\left\{\left(\hat{\upsilon}_{\text{N}}(\tau; \tilde{\boldsymbol{Z}}_{\pi}), \hat{\upsilon}_{\text{N}}(\tau; \tilde{\boldsymbol{Z}}_{\pi'})\right) : \tau \in \mathscr{T}\right\} \tag{I.1}$$

converges weakly to a tight process $\left\{\left(\upsilon(\tau), \upsilon'(\tau)\right) : \tau \in \mathscr{T}\right\}$, where $\left(\upsilon(\cdot), \upsilon'(\cdot)\right)$ are independent Brownian bridges. Observe that assumption A.1 in the main text implies that the inverse map is Hadamard-differentiable by Van der Vaart and Wellner (1996, Lemma 3.9.23) and that $\sup|\hat{\varphi}(\tau)| < \infty$. Therefore, it follows by Chung and Olivares (2021, Theorem 2) and the Delta-method that the process (I.1) converges weakly to a tight process $\left\{\left(\upsilon(\tau), \upsilon'(\tau)\right) : \tau \in \mathscr{T}\right\}$, where $\left(\upsilon(\cdot), \upsilon'(\cdot)\right)$ are independent Brownian bridges, as desired. This finishes the proof.

I.2 Proof of Theorem 2

We seek to show the asymptotic behavior of $R_{\text{N}}^{\tilde{K}}$, the permutation distribution based on the two-sample martingale-transformed quantile process $\tilde{\upsilon}_{\text{N}}(\tau; \tilde{\boldsymbol{Z}})$, where we have used the numerical equivalence as we argued in the proof of Theorem 1. Since $\tilde{\upsilon}_{\text{N}}(\tau; \tilde{\boldsymbol{Z}})$ is a continuous mapping by the arguments in the proof of Lemma 2, then it suffices by the continuous mapping theorem (CMT) for randomization distributions (Chung and Romano, 2016, Lemma A.6) to establish the asymptotic behavior of the permutation distribution based on $\tilde{\upsilon}_{\text{N}}(\tau; \tilde{\boldsymbol{Z}})$, given by

$$\tilde{\upsilon}_{\text{N}}(\tau; \tilde{\boldsymbol{Z}}) = \hat{\upsilon}_{\text{N}}(\tau; \tilde{\boldsymbol{Z}}) - \psi_g\left(\hat{\upsilon}_{\text{N}}\right)(\tau; \tilde{\boldsymbol{Z}}) \ .$$

Thus, it suffices to prove by Lehmann and Romano (2022, Theorem 17.2.3) that

$$\left\{\left(\hat{\upsilon}_{\text{N}}(\tau; \tilde{\boldsymbol{Z}}_{\pi}) - \psi_g(\hat{\upsilon}_{\text{N}})(\tau; \tilde{\boldsymbol{Z}}_{\pi}), \hat{\upsilon}_{\text{N}}(\tau; \tilde{\boldsymbol{Z}}_{\pi'}) - \psi_g(\hat{\upsilon}_{\text{N}})(\tau; \tilde{\boldsymbol{Z}}_{\pi'})\right) : \tau \in \mathscr{T}\right\} \tag{I.2}$$

converges weakly to a tight process $\left\{\left(\zeta(\tau), \zeta'(\tau)\right) : \tau \in \mathscr{T}\right\}$, where $(\zeta(\tau), \zeta'(\tau))$ denotes a vector of two independent Brownian motion processes given by $\upsilon(\tau) - \varphi(\upsilon)(\tau)$, where $\upsilon(\cdot)$ is a tight Brownian bridge (and the same is true if we replace $\upsilon(\tau)$ by $\upsilon'(\tau)$).

We proved in Theorem 1 that (I.1) converges weakly to $\left\{\left(\upsilon(\tau), \upsilon'(\tau)\right) : \tau \in \mathscr{T}\right\}$, where $\left(\upsilon(\cdot), \upsilon'(\cdot)\right)$ are independent Brownian bridges. Furthermore, the continuity of $\psi_g(\cdot)$ implies that $\left\{\left(\psi_g(\hat{\upsilon}_{\mathrm{N}})(\tau; \tilde{\boldsymbol{Z}}_\pi), \psi_g(\hat{\upsilon}_{\mathrm{N}})(\tau; \tilde{\boldsymbol{Z}}_{\pi'})\right) : \tau \in \mathscr{T}\right\}$ converges weakly to $\left(\psi_g(\upsilon), \psi_g(\upsilon')\right)(\cdot)$ by the CMT for randomization distributions (Chung and Romano, 2016, Lemma A.6). Here, continuity follows by noting ψ_g is a Fredholm operator on a Banach space, hence a bounded operator. But an operator between normed spaces is bounded if and only if it is a continuous operator (Abramovich and Aliprantis, 2002). Then, the weak limit of (I.2) follows by Slutsky's theorem for randomization distributions (Chung and Romano, 2013, Theorem 5.2). This finishes the proof of our claim and the first part of the theorem.

For the second part of the theorem, we note that the distribution of $\tilde{K}$, *i.e.*, the distribution of the norm of a tight Brownian motion process, is strictly increasing and absolutely continuous with a positive density (Beran and Millar, 1986, Proposition 2). Thus, under the conditions of the theorem, $\hat{r}_{\mathrm{N}}(1-\alpha) \xrightarrow{\mathrm{p}} r(1-\alpha) = \inf\{t : H(t) \geq 1-\alpha\}$ by Lehmann and Romano (2022, Problem 11.30), concluding the proof.

I.3 Proof of Corollary 1

We have that $\Pr\left\{\tilde{K}_{\mathrm{N}} > \hat{r}_{\mathrm{N}}(1-\alpha)\right\} \leq \mathbb{E}\left[\phi(Z)\right] \leq \Pr\left\{\tilde{K}_{\mathrm{N}} \geq \hat{r}_{\mathrm{N}}(1-\alpha)\right\}$ from the construction of the permutation test based on $\tilde{K}_{\mathrm{N}}(\boldsymbol{Z})$. Hence, Theorem 2 implies $\mathbb{E}\left[\phi(Z)\right] \to 1 - H(r(1-\alpha) = \alpha$, as desired.

II Proofs of Lemmas

II.1 Proof of Lemma 1

We are interested in showing the asymptotic behavior of the test statistic under the null hypothesis. We begin by rewriting (6) in the main text as

$$\hat{\upsilon}_{\mathrm{N}}(\tau;\boldsymbol{Z}) = \sqrt{\frac{mn}{N}}\hat{\varphi}(\tau)\left\{\hat{\gamma}(\tau)-\gamma(\tau)\right\} - \sqrt{\frac{mn}{N}}\hat{\varphi}(\tau)\left\{\hat{\gamma}-\gamma\right\} + \sqrt{\frac{mn}{N}}\hat{\varphi}(\tau)\left\{\gamma(\tau)-\gamma\right\} \ , \tag{II.1}$$

where the last term in (II.1) is zero under the null hypothesis. Develop further to obtain

$$\begin{aligned}\hat{\upsilon}_{\mathrm{N}}(\tau;\boldsymbol{Z}) &= \sqrt{\frac{mn}{N}}\varphi(\tau)\left\{\hat{\gamma}(\tau)-\gamma(\tau)\right\} - \sqrt{\frac{mn}{N}}\varphi(\tau)\left\{\hat{\gamma}-\gamma\right\} \\ &\quad + \sqrt{\frac{mn}{N}}\left[\hat{\varphi}(\tau)-\varphi(\tau)\right]\left\{\hat{\gamma}(\tau)-\gamma(\tau)\right\} - \sqrt{\frac{mn}{N}}\left[\hat{\varphi}(\tau)-\varphi(\tau)\right]\left\{\hat{\gamma}-\gamma\right\} \\ &= \sqrt{\frac{mn}{N}}\varphi(\tau)\left\{\hat{\gamma}(\tau)-\gamma(\tau)\right\} - \sqrt{\frac{mn}{N}}\varphi(\tau)\left\{\hat{\gamma}-\gamma\right\} + o_p(1) \ , \end{aligned} \tag{II.2}$$

where the $o_p(1)$ term holds uniformly over $\mathscr{T}$ by Assumption A.3 (ii). For the sake of notational compactness, denote $\upsilon_{\mathrm{N}}(\tau;\boldsymbol{Z})$ and $\xi_{\mathrm{N}}(\tau;\boldsymbol{Z})$ as the first and second summands in (II.2), respectively. Under Assumptions A.1 and A.2 in the main text, $\{\upsilon_{\mathrm{N}}(\tau;\boldsymbol{Z}) : \tau \in \mathscr{T}\}$ converges weakly in $\ell^\infty(\mathscr{T})$ to a Brownian bridge process $\upsilon(\cdot)$ by Shorack and Wellner (2009, Theorem 2, Ch. 18).

By Assumption A.3 (i), the second term on the right-hand side of (II.2) is in $\ell^\infty(\mathscr{T})$ if and only if $\sup_{\mathscr{T}}|\varphi(\tau)| < \infty$. But this follows by A.1, which implies F_0 is Lipschitz continuous and therefore $\sup_{\mathscr{T}}|\varphi(\tau)| < \infty$. Therefore, $\{\xi_{\mathrm{N}}(\tau;\boldsymbol{Z}) : \tau \in \mathscr{T}\}$ converges weakly in $\ell^\infty(\mathscr{T})$ to a mean zero Gaussian process $\xi(\cdot)$ with covariance function $\mathbb{C}(\xi(\tau_1),\xi(\tau_2)) = \varphi(\tau_1)\varphi(\tau_2)\sigma_0^2$. Thus, the limit process $\upsilon(\cdot)+\xi(\cdot)$ is a Gaussian process with zero mean and covariance function

$$\mathbb{C}(\upsilon(\tau_1), \xi(\tau_2)) = \min\{\tau_1, \tau_2\} - \tau_1\tau_2 + \varphi(\tau_1)\varphi(\tau_2)\sigma_0^2 \\ + \tau_1(1-\tau_1)\varphi(\tau_2)\{\mathbb{E}(Y_0|Y_0 \le F_0^{-1}(\tau_2)) - \mathbb{E}(Y_0|Y_0 > F_0^{-1}(\tau_2))\} \ . \quad \text{(II.3)}$$

This finishes the first part of the proof. Note that the maps $v \to \|v\|$ from $\ell^\infty(\mathscr{T})$ into $\mathbb{R}$ are continuous with respect to the supremum norm. Then, a direct application of the continuous mapping theorem (CMT) yields the final result. This finishes the proof.

II.2 Proof of Lemma 2

Recall the martingale-transformation of the quantile process, eq. (14) in the main text, is given by

$$\begin{aligned} \tilde{\upsilon}_{\mathrm{N}}(\tau; \boldsymbol{Z}) &= \hat{\upsilon}_{\mathrm{N}}(\tau; \boldsymbol{Z}) - \psi_g\left(\hat{\upsilon}_{\mathrm{N}}\right)(\tau; \boldsymbol{Z}) \\ &= \upsilon_{\mathrm{N}}(\tau; \boldsymbol{Z}) + \xi_{\mathrm{N}}(\tau; \boldsymbol{Z}) - \psi_g\left(\hat{\upsilon}_{\mathrm{N}}\right)(\tau; \boldsymbol{Z}) + o_p(1) \ , \end{aligned} \quad \text{(II.4)}$$

where the second equality follows by the asymptotic expansion (II.2) and the $o_p(1)$ term holds uniformly over $\mathscr{T}$. By properties of the compensator ψ_g and (II.2), we have that

$$\psi_g\left(\hat{\upsilon}_{\mathrm{N}}\right)(\tau; \boldsymbol{Z}) = \psi_g\left(\upsilon_{\mathrm{N}}\right)(\tau; \boldsymbol{Z}) + \xi_{\mathrm{N}}(\boldsymbol{Z}) + o_p(1) \ . \quad \text{(II.5)}$$

Plugging (II.5) into (II.4) yields

$$\tilde{\upsilon}_{\mathrm{N}}(\tau; \boldsymbol{Z}) = \upsilon_{\mathrm{N}}(\tau; \boldsymbol{Z}) - \psi_g\left(\upsilon_{\mathrm{N}}\right)(\tau; \boldsymbol{Z}) + o_p(1) \ , \quad \text{(II.6)}$$

and therefore, $\{\tilde{\upsilon}_{\mathrm{N}}(\tau; \boldsymbol{Z}) : \tau \in \mathscr{T}\}$ converges weakly to $\zeta(\cdot)$, the standard Brownian motion by Khmaladze (1981, 4.3). This finishes the proof of the first part of the lemma. A direct application of the CMT as in the proof of Lemma 1 finishes the proof.

III Additional Simulation Results

In this Appendix, we compare our approach, based on the quantile process, with that of Chung and Olivares (2021, **CO21** hereafter), based on CDF comparisons via simulations. The numerical results in Table 1 focus on size control since this is one of the permutation tests' salient features. To help us compare both methods, we borrowed the results from **CO21** and stick to their exact design.

Table 1: Size of $\alpha = 0.05$ tests H_0 : Constant Treatment Effect ($\gamma = 1$).

		Distributions		
N	Method	Normal	Lognormal	Student's t
$N = 200$	CO21	0.0236	0.0354	0.0428
$n = 120$, $m = 80$	mtqPermTest	0.0502	0.0490	0.0435
$N = 800$	CO21	0.0288	0.0470	0.0438
$n = 500$, $m = 300$	mtqPermTest	0.0484	0.0490	0.0526
$N = 1000$	CO21	0.0292	0.0480	0.0474
$n = 600$, $m = 400$	mtqPermTest	0.0540	0.0464	0.0456

We calculate the empirical rejection frequencies based on 5000 simulations. C&O results extracted from Chung and Olivares (2021, Table 1).

Overall, we see that both tests deliver similar rejection frequencies under the null and have small size distortions across specifications. However, two observations are in place. First, **CO21** test is generally more conservative relative to **mtPermTest**, particularly in the Normal case. Second, **mtPermTest** displays a remarkable finite sample performance even using a fairly small sample size ($N = 200$). Therefore, we conclude from the numerical evidence that the quantile-based permutation test offers an advantage with respect to the CDF case.

IV BGH and The Durbin Problem

BGH apply a Fisher-randomization test using the plug-in method for the individual hypotheses (20) (Bitler, Gelbach, and Hoynes, 2017, Section V.B, p. 694). To formalize the ongoing discussion, we need more notation. Let the observed data for each mutually exclusive subgroup be given by $\boldsymbol{Z}^s = \left(Z_1^s \ldots, Z_{N_s}^s\right) = \left(Y_{s,1}^1, \ldots, Y_{s,m_s}^1, Y_{s,1}^0, \ldots, Y_{s,n_s}^0\right)$ for all $1 \leq s \leq \mathcal{S}$, where every subgroup $\boldsymbol{Z}^s$ has $N_s = m_s + n_s$ observations.

Under the assumptions of the CSTE model, $Y_{1,s} = Y_{0,s} + \delta_s$ for all s. Then, one might shift the observations in the control group, $Y_{0,s}$, by adding these subgroup-specific δ_s, that we can estimate as the ATEs within subgroup s, to the actual outcome. Let $\hat{Y}_{0,s} = Y_{0,s} + \hat{\delta}_s$ be the simulated outcomes within subgroup s. This simple transformation across subgroups yields a simulated outcome under treatment, $\{\hat{Y}_{0,s} : 1 \leq s \leq \mathcal{S}\}$.

BGH argue that if the CSTE model is a good representation of HTE, then the distribution of the simulated outcomes should be "close" in some sense to $F_{1,s}(\cdot)$, the distribution of the observed outcomes under treatment. Therefore, BGH's permutation test is based on the two-sample Kolmogorov–Smirnov test statistic (2SKS) for each subgroup,

$$K^s_{\mathrm{N},\hat{\delta}}(\boldsymbol{Z}^s) = \sup_{y \in \mathbb{R}} \left| V^s_{\mathrm{N}}(y, \hat{\delta}_s; \boldsymbol{Z}^s) \right| , \tag{IV.1}$$

where

$$V^s_{\mathrm{N}}(y, \hat{\delta}_s; \boldsymbol{Z}^s) = \sqrt{\frac{m_s n_s}{N_s}} \left\{ \hat{F}_{1,s}(y) - \hat{F}_{0,s}(y - \hat{\delta}_s) \right\} . \tag{IV.2}$$

From the previous test statistic, one can define BGH's permutation test as follows. Let $\boldsymbol{G}_s$ be the set of all permutations π of $\{1, \ldots, N_s\}$. For a fixed subgroup s, BGH's permutation test based on 2SKS rejects the individual hypothesis (20) if the observed (IV.1) exceeds the $1 - \alpha/\mathcal{S}$ quantile of the permutation distribution:

$$\hat{R}^{K(\hat{\delta})}_{\mathrm{N},s}(t) = \frac{1}{N_s!} \sum_{\pi \in \mathbf{G}_s} \mathbb{1}\left\{ K^s_{\mathrm{N},\hat{\delta}}\left(Z^s_{\pi(1)}, \ldots, Z^s_{\pi(N_s)}\right) \leq t \right\} . \tag{IV.3}$$

IV.1 Commentary on BGH's Approach

BGH state that the critical values derived from (IV.3) are asymptotically valid even in the presence of estimated parameters (Section V, p. 694). However, this claim is false. To establish the asymptotic validity of the permutation test, we need to show that the permutation distribution (IV.3) approximates the true unconditional sampling distribution of (IV.1). Under relatively weak assumptions, we can show that (IV.3) behaves like the distribution of (the supremum of) a Brownian bridge (Chung and Olivares, 2021, Theorem 2). In contrast, the limiting distribution of the 2SKS is given by the distribution of (the supremum of) a different Gaussian process; it has mean 0 and a covariance structure that depends on unknown parameters as proved in Ding, Feller, and Miratrix (2016, Theorem 4). Therefore, the permutation test based on the 2SKS fails to control the type 1 error asymptotically.

Then, how do we make sense of BGH's claims? The authors' justification relies on a result by Præstgaard (1995) that states the permutation empirical process converges weakly to a Brownian bridge corresponding to a mixture measure. Indeed, the testing problem in BGH's environment satisfies the premises in Præstgaard's (1995), so (IV.3) asymptotically does behave like the process in Præstgaard (1995).

In other words, the asymptotic behavior of the permutation distribution (IV.3) does not change when the estimated δ_s enters the test statistic instead of the known value δ_s. However, it is not the case for the true unconditional limiting distribution of the test statistic. The asymptotic behavior of the 2SKS statistic does change in the presence of estimated parameters. Therefore, the permutation distribution does not mimic the true unconditional distribution of the test statistic in large samples when δ_s is being estimated, invalidating BGH's permutation test. Will BGH's approach ever be valid? Yes, but only in the infeasible scenario when we know the subgroup-specific δ_s. In fact, in this case, the permutation test achieves the finite sample exactness. Therefore, BGH's claim that their method yields asymptotically valid inference is only well-grounded in this extraordinary case and incorrect otherwise.

V Empirical Results from BGH

Table 2 displays the results from BGH empirical analysis and the proposed permutation test based on the quantile process in Section 4.2. Columns 3–4 contain the empirical results from BGH's Table 2, which we include verbatim for a fair comparison. Meanwhile, columns 5–6 contain the results from our proposed method. We note that a joint test of the family of null hypotheses across these subgroups rejects if we reject any one of the subgroup-specific null hypotheses (see BGH's Section V).

Table 2: Testing for Heterogeneity in the Treatment Effect by Subgroups, Time-varying mean treatment effects by subgroup with participation adjustment

Subgroup	Number of Tests	BGH's Permutation Test Number of Reject at 10%	BGH's Permutation Test Number of Reject at 5%	Asymptotically Valid Permutation Test Number of Reject at 10%	Asymptotically Valid Permutation Test Number of Reject at 5%
Full Sample	7	4	4	3	3
Education	21	3	1	3	1
Age of youngest child	21	3	1	3	3
Marital status	21	2	1	3	2
Earnings level seventh Q pre-RA	21	2	1	0	0
Number of pre-RA Q with earnings	21	1	0	1	0
Welfare receipt seventh Q pre-RA	14	3	3	2	2
Education subgroups interacted with					
Age of youngest child	49	1	0	6	5
Marital status	35	3	3	4	2
Earnings level seventh Q pre-RA	63	1	0	0	0
Number of pre-RA Q with earnings	63	0	0	2	1
Welfare receipt seventh Q pre-RA	42	1	0	1	0
Age of youngest child interacted with					
Marital status	35	1	1	3	1
Earnings level seventh Q pre-RA	63	0	0	1	0
Number of pre-RA Q with earnings	49	1	1	3	1
Welfare receipt seventh Q pre-RA	42	1	0	1	0
Marital status subgroup interacted with					
Earnings level seventh Q pre-RA	63	2	1	1	0
Number of pre-RA Q with earnings	63	0	0	2	0
Welfare receipt seventh Q pre-RA	42	1	0	2	1
Earnings level seventh Q pre-RA subgroups interacted with					
Number of pre-RA Q with earnings	49	0	0	2	0
Welfare receipt seventh Q pre-RA	42	1	1	1	1
Number of quarters any earnings pre-RA subgroup interacted with					
Welfare receipt seventh Q pre-RA	42	0	0	2	1

All reported results account for multiple testing using Bonferroni adjustment. We use 1000 permutations for the stochastic approximation of the permutation distribution.

VI Numerical Computation of $\tilde{v}_{\mathrm{N}}(\tau; \boldsymbol{Z})$

This section illustrates how to numerically compute the martingale-transformed test statistic. It largely mirrors Sections 3.3 and 3.4 in CO21. We include this section to keep things self-contained. We recommend the interested reader see Bai (2003) and Parker

(2013) as well. In this section and only to convey the main idea in the most familiar way, y and x act as surrogates for the "outcome" and "regressor" in the usual linear regression model. Thus, the reader should not confuse them with the outcome or covariates in the main text.

Consider $\tilde{v}_{\mathrm{N}}(\tau; \boldsymbol{Z})$ defined in (14) with τ taking discrete values, thus replacing integral with sums. For instance, suppose $\varepsilon = t_0 < t_1 < \cdots < t_q < \tau_{q+1} = 1 - \varepsilon$ is a partition of $\mathscr{T}$ and that τ takes on values on $t_1, \ldots, t_q$. Write $\tilde{v}_{\mathrm{N}}(t; \boldsymbol{Z})$ in differentiation form:

$$d\tilde{v}_{\mathrm{N}}(t; \boldsymbol{Z}) = d\,\hat{v}_{\mathrm{N}}(t; \boldsymbol{Z}) - \dot{g}(t)' C(t)^{-1} \int_t^1 \dot{g}(r) d\,\hat{v}_{\mathrm{N}}(r; \boldsymbol{Z}) dt \ . \tag{VI.1}$$

Define $dt_i = t_{i+1} - t_i$, and let

$$\begin{aligned} y_i &= d\,\hat{v}_{\mathrm{N}}(t_i; \boldsymbol{Z}) \\ x_i &= \dot{g}(t_i) dt_i \\ C(t_i) &= \sum_{k=i}^{q+1} x_k x_k' \\ \int_t^1 \dot{g}(r) d\,\hat{v}_{\mathrm{N}}(r; \boldsymbol{Z}) &= \sum_{k=i}^{q+1} x_k y_k \ , \end{aligned}$$

then the right hand side of (VI.1) can be interpreted as the recursive residuals:

$$y_i - x_i' \left(\sum_{k=i}^{q+1} x_k x_k' \right)^{-1} \sum_{k=i}^{q+1} x_k y_k = y_i - x_i' \hat{\boldsymbol{\beta}}_i \ , \tag{VI.2}$$

where $\hat{\boldsymbol{\beta}}_i$ is the OLS estimator based on the last $q - i + 2$ observations. From here, it is straightforward to see that the cumulative sum (integration from $[\varepsilon, t_i)$) of (VI.2) gives rise to a Brownian motion process. Note that here τ takes discrete values but, in the continuous case, the previous construction boils down to the conclusion in Lemma 1.

With this in mind, we now turn to the numerical calculation of the compensator and the transformed statistic, $\tilde{v}_{\mathrm{N}}(\tau; \boldsymbol{Z})$. For the sake of exposition, suppose that the quantile density and score functions are known (the construction is exactly analogous if we replace $\dot{g}$ with an estimate $\dot{g}_{\mathrm{N}}$ satisfying assumption A.4 in the main text). Observe that the computation of the compensator involves numerical integration. Assume the partition $\{t_i\}$ is evenly spaced, with the accuracy depending on how fine the grid is, *i.e.*, the number of points q. Stack y_i and x_i in the following manner

$$\mathbf{X}_i = \sqrt{\frac{1}{q}} \begin{pmatrix} \dot{g}_1(t_{q+1}) & \dot{g}_2(t_{q+1}) \\ \dot{g}_1(t_q) & \dot{g}_2(t_q) \\ \vdots & \vdots \\ \dot{g}_1(t_i) & \dot{g}_2(t_i) \end{pmatrix}, \quad \mathbf{y}_i = \sqrt{q} \begin{pmatrix} \hat{\upsilon}_{\mathrm{N}}(t_{q+1}; \boldsymbol{Z}) & - & \hat{\upsilon}_{\mathrm{N}}(t_q; \boldsymbol{Z}) \\ \hat{\upsilon}_{\mathrm{N}}(t_q; \boldsymbol{Z}) & - & \hat{\upsilon}_{\mathrm{N}}(t_{q-1}; \boldsymbol{Z}) \\ & \vdots & \\ \hat{\upsilon}_{\mathrm{N}}(t_i; \boldsymbol{Z}) & - & \hat{\upsilon}_{\mathrm{N}}(t_{i-1}; \boldsymbol{Z}) \end{pmatrix},$$

where $\dot{g}_1(s) = 1$ and $\dot{g}_2(s) = \dot{f}_0(F_0^{-1}(s))/f_0(F_0^{-1}(s))$. The least squares estimate based on the last $q-i+2$ observations described on the right-hand side of (VI.2) can be written as $\hat{\boldsymbol{\beta}}_i = (\mathbf{X}_i'\mathbf{X}_i)^{-1} \mathbf{X}_i'\mathbf{y}_i$, which implies that the two-sample martingale-transformed quantile process (14) can be obtained by numerically integrating from $[\varepsilon, t_i)$, *i.e.*,

$$\tilde{\upsilon}_{\mathrm{N}}(t_i; \boldsymbol{Z}) = \hat{\upsilon}_{\mathrm{N}}(t_i; \boldsymbol{Z}) - \frac{1}{q} \sum_{j=1}^{i} x_j' \hat{\boldsymbol{\beta}}_j \ .$$

Then, we can calculate the martingale-transformed test statistic, $\tilde{K}_{\mathrm{N}}(\boldsymbol{Z})$, as

$$\max_{1 \le i \le q} \left| \hat{\upsilon}_{\mathrm{N}}(t_i; \boldsymbol{Z}) - \frac{1}{q} \sum_{j=1}^{i} x_j' \hat{\boldsymbol{\beta}}_j \right| \ .$$

References

Abramovich, Y. A. and Aliprantis, C. D. (2002). *An Invitation to Operator Theory*, volume 1. American Mathematical Soc.

Bai, J. (2003). Testing parametric conditional distributions of dynamic models. *Review of Economics and Statistics*, 85(3):531–549.

Beran, R. and Millar, P. (1986). Confidence sets for a multivariate distribution. *The Annals of Statistics*, pages 431–443.

Bitler, M. P., Gelbach, J. B., and Hoynes, H. W. (2017). Can variation in subgroups' average treatment effects explain treatment effect heterogeneity? evidence from a social experiment. *Review of Economics and Statistics*, 99(4):683–697.

Chung, E. and Olivares, M. (2021). Permutation test for heterogeneous treatment effects with a nuisance parameter. *Journal of Econometrics*, 225(2):148–174.

Chung, E. and Romano, J. P. (2013). Exact and asymptotically robust permutation tests. *The Annals of Statistics*, 41(2):484–507.

Chung, E. and Romano, J. P. (2016). Multivariate and multiple permutation tests. *Journal of Econometrics*, 193(1):76–91.

Ding, P., Feller, A., and Miratrix, L. (2016). Randomization inference for treatment effect variation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Khmaladze, E. V. (1981). Martingale approach in the theory of goodness-of-fit tests. *Theory of Probability & Its Applications*, 26(2):240–257.

Lehmann, E. L. and Romano, J. P. (2022). *Testing statistical hypotheses*. Springer.

Parker, T. (2013). A comparison of alternative approaches to supremum-norm goodness-of-fit tests with estimated parameters. *Econometric Theory*, 29(05):969–1008.

Præstgaard, J. T. (1995). Permutation and bootstrap kolmogorov-smirnov tests for the equality of two distributions. *Scandinavian Journal of Statistics*, pages 305–322.

Shorack, G. R. and Wellner, J. A. (2009). *Empirical processes with applications to statistics*. SIAM.

Van der Vaart, A. W. and Wellner, J. (1996). *Weak convergence and empirical processes: with applications to statistics*. Springer Science & Business Media.