Supplement for "Can Variation in Subgroups's Average Treatment Effects Explain Treatment Effect Heterogeneity? Evidence from a Social Experiment: A Comment"[*]

EunYi Chung[†] Mauricio Olivares[‡]

University of Illinois at Urbana-Champaign

April 2, 2020

Abstract

This document provides additional material for the authors' paper "Can Variation in Subgroups's Average Treatment Effects Explain Treatment Effect Heterogeneity? Evidence from a Social Experiment: A Comment". It includes a formal description of the statistical environment, the general construction of a permutation test, and additional empirical results. We also include a brief description of the covariance structures needed for the numerical exercise, as well as their analytical expressions for a specific DGP.

Keywords: Permutation Test, Khmaladze transformation, Heterogeneous treatment effects, Connecticut's Jobs First.

JEL Classification: C12, C14, C46.

[*]We are grateful to MDRC for granting access to the experimental data used in this paper. We would also like to thank Marianne Bitler, Jonah Gelbach, and Hilary Hoynes for their useful comments and generous feedback that led to the improvement of the paper as well as for making their code for replication publicly available. Lastly, we thank Dan Bernhardt for helpful discussion and encouragement during this project. All errors are our own. Address correspondence to EunYi Chung (eunyi@illinois.edu).

[†]EunYi Chung is an Assistant Professor in the Department of Economics, University of Illinois at Urbana-Champaign. Email: eunyi@illinois.edu

[‡]Mauricio Olivares is a Ph.D. student in the Department of Economics, University of Illinois at Urbana-Champaign. Email: lvrsgnz2@illinois.edu

1 Statistical Environment

For our purposes we cast BGH's testing problem under the umbrella of the potential outcomes framework (Neyman, 1990; Rubin, 1974). This allows for greater simplicity of exposition while capturing all the features of the testing problem at hand. Let $Y_i \in \mathbb{R}$ be the observed outcome of interest for the ith unit, $D_i \in \{0,1\}$ is the treatment indicator for whether the unit receives treatment or not. In addition, let X_i denote the observed, baseline covariates for the ith unit. Denote $Y_i(1)$ the potential outcome for the ith unit if treated ($D_1 = 1$), and similarly, $Y_i(0)$ represents the potential outcome for the ith unit if not treated ($D_i = 0$).

The treatment effect is defined as the difference between the potential outcomes, $\delta_i = Y_i(1) - Y_i(0)$ for $1 \leq i \leq N$. The treatment effect is *constant* if $\delta_i = \delta$ for all i, otherwise we say the treatment effect is *heterogeneous*. Thus if the treatment effect is constant then $Y_i(1) = Y_i(0) + \delta$, *i.e.* the potential outcomes are a constant shift apart.

Throughout we assume that the mutually exclusive subgroups are formed from observed covariates, and are taken as given. Denote $\mathcal{J}$ the total number of such subgroups.[1] Let $Y_i^j(D)$ be the potential outcome of unit i under assignment D, given that it is a member of subgroup j for $1 \leq j \leq \mathcal{J}$. Therefore, the treatment effect for the ith unit in subgroup j is given by $\delta_i^j = Y_i^j(1) - Y_i^j(0)$ for all mutually exclusive subgroups $j \in \{1, \ldots, \mathcal{J}\}$.

1.1 Null Hypotheses

The main idea behind the strategy proposed by BGH rests upon the simulated outcomes under treatment, given by $Y_i^{*j}(1) = Y_i^j(0) + \delta^j$. If the CTE model is correct, then $Y_i^{*j}(1) = Y_i(1)$ for all $1 \leq i \leq N$, and $1 \leq j \leq \mathcal{J}$. Then, the null hypothesis of interest is

$$H_0^s : Y_i^j(0) + \delta^j = Y_i^j(1) \quad \text{for some} \quad \delta^j, \quad \text{for all mutually exclusive } 1 \leq j \leq \mathcal{J} \ . \tag{1}$$

If δ^j were to be known, then (1) becomes a sharp null. Hypotheses like (1) are, however, not directly testable—we only observe at most one potential outcome for each individual. A different but testable hypothesis is available if we consider the marginal distributions of earnings for individuals who were treated *vis-à-vis* women who were not.

More formally, let $F_1^{*j}(y)$ and $F_1^j(y)$ be the CDFs of the simulated and actual treatment group earnings distributions, for $1 \leq j \leq \mathcal{J}$. The testable hypothesis becomes the joint null hypothesis

$$\mathbf{H}_0 : F_1^{*j}(y) = F_0^j(y) \ , \text{ for all mutually exclusive } j \in \{1, \ldots, \mathcal{J}\} \, , \text{ for some } \delta^j \ , \tag{2}$$

[1] We may think of subgroups as constructed from the observed covariates X_i by the function $S : \text{supp}(X_i) \to \mathcal{S}$, where $\mathcal{S}$ is a finite set of cardinality $\mathcal{J}$.

The testing problem defined in (2) can be seen as a multiple testing problem in which every individual hypothesis j, given by

$$H_{0,j} : F_1^{*j}(y) = F_1^j(y) \ , \text{ for some } \delta^j \ , \tag{3}$$

characterizes the CTE model since (a) effects vary across subgroups, but (b) are the same for each member of the subgroup. Thus, to reject the null hypothesis that all these distributions are the same for each subgroup—hypothesis (2)—it suffices that we reject the individual null hypothesis (3) for at least one of the subgroups.[2].

1.2 Test Statistic and Simulated Outcomes

For $1 \leq j \leq \mathcal{J}$, let n_j and m_j be the number of observations in the control and treatment group in subgroup j, with $n_j + m_j = N_j$. Let

$$Z^j = (Z_1^j, \ldots, Z_{N_j}^j) = (Y_1^j(1), \ldots, Y_{m_j}^j(1), Y_1^j(0), \ldots, Y_{n_j}^j(0))$$

be the pooled data from the treatment and control groups for each subgroup j.

BGH, propose a test statistic for null hypotheses like (3) based on the comparison of the empirical CDFs

$$\hat{F}_1^{*j}(y) = \frac{1}{n_j} \sum_{i=1}^{n_j} \mathbb{1}_{\{\hat{Y}_i^{*j} \leq y\}} \quad \text{and} \quad \hat{F}_1(y) = \frac{1}{m_j} \sum_{i=1}^{m_j} \mathbb{1}_{\{Y_i^j(1) \leq y\}} \ ,$$

where $\hat{Y}_i^{*j}$ is an estimate of the simulated earnings. Under the CTE model, we have that δ^j is identified and $\sqrt{N_j}$-consistently estimable as the difference in sample means from treatment and control groups, for all subgroups $1 \leq j \leq \mathcal{J}$. Then, $\hat{Y}_i^{*j}$ is given by $\hat{Y}_i^{*j} = Y_i^j(0) + \hat{\delta}^j$.

This gives rise to the two-sample Kolmogorov–Smirnov statistic for subgroup j as

$$K_{m,n,\hat{\delta}}^j(Z^j) = \sup_y \left| V_{m,n}^j(y, \hat{\delta}^j) \right| \ , \tag{4}$$

where

$$V_{m,n}^j(y, \hat{\delta}^j) = \sqrt{\frac{m_j n_j}{N_j}} \left(\hat{F}_1^{*j}(y) - \hat{F}_1^j(y) \right) \tag{5}$$

is the two-sample empirical process based on the simulated and actual treatment group earnings distributions for that subgroup.

[2]Naively testing for treatment effect variation for each subgroup at level α may lead us to flawed inference though. With such a procedure the probability of one or more false rejections rapidly increases with the number of subgroups. To put it in other words, the probability of falsely claiming that the treatment effect is heterogeneous for some subgroup may be greater than α.

1.3 Permutation Test

BHG propose a permutation test to conduct inference in this setting. In particular, they propose a permutation test for each individual hypothesis (3), and then testing the joint null hypothesis (2) if any of the individual nulls is rejected. To control for the family-wise error rate, BGH use a Bonferroni adjustment.

In this section we present the construction of a permutation test for the individual hypotheses (3), for a fixed subgroup $j \in \{1, \dots, \mathcal{J}\}$, and a Bonferroni correction[3]. The construction of such a test, however, requires some additional notation. Let $\mathbf{G}_j$ be the set of all permutations π of $\{1, \dots, N_j\}$, with cardinality $M = N_j!$. For each $\pi \in \mathbf{G}_j$, generate the simulated earnings under treatment by adding $\hat{\delta}^j$, and collect into

$$\hat{Z}^j_\pi = \left(Z^j_{\pi(1)}, \dots, Z^j_{\pi(m_j)}, Z^j_{\pi(m_j+1)} + \hat{\delta}^j, \dots, Z^j_{\pi(m_j+n_j)} + \hat{\delta}^j\right)$$

Observe that we create the simulated treatment group earnings by adding $\hat{\delta}^j$ to the permutation-generated control group. Without loss of generality, assume π_1 is the identity permutation. Then the observed Kolmogorov–Smirnov statistic, calculated based on the original data, is given by

$$K_{m,n,\hat{\delta}}(\hat{Z}^j) = K_{m,n,\hat{\delta}}(\hat{Z}^j_{\pi_1}) \ .$$

Given $\hat{Z}^j = \hat{Z}^j_{\pi_1} = z$, recompute $K_{m,n,\hat{\delta}}(z)$ for all permutations $\pi \in \mathbf{G}_j$ and denote by

$$K^{(1)}_{m,n,\hat{\delta}}(z) \le K^{(2)}_{m,n,\hat{\delta}}(z) \le \cdots \le K^{(M)}_{m,n,\hat{\delta}}(z) \ ,$$

the ordered values of $\{K_{m,n,\hat{\delta}}(z_\pi) : \pi \in \mathbf{G}_j\}$, where z_π denotes the action of $\pi \in \mathbf{G}_j$ on $z \in \mathbb{R}$. Let $k = M - \lfloor M\,\alpha/\mathcal{J} \rfloor$ and define

$$\begin{aligned} M^+(z) &= \left|\{1 \le l \le M : K^{(l)}_{m,n,\hat{\delta}}(z) > K^{(k)}_{m,n,\hat{\delta}}(z)\}\right| \\ M^0(z) &= \left|\{1 \le l \le M : K^{(l)}_{m,n,\hat{\delta}}(z) = K^{(k)}_{m,n,\hat{\delta}}(z)\}\right| \ . \end{aligned}$$

Using this notation, the permutation test for subgroup j is given by

$$\phi^j(z) = \begin{cases} 1 & K_{m,n,\hat{\delta}}(z) > K^{(k)}_{m,n,\hat{\delta}}(z) \\ a(z) & K_{m,n,\hat{\delta}}(z) = K^{(k)}_{m,n,\hat{\delta}}(z) \\ 0 & K_{m,n,\hat{\delta}}(z) < K^{(k)}_{m,n,\hat{\delta}}(z) \end{cases} \ , \tag{6}$$

where

$$a(z) = \frac{M\,\alpha - M^+(z)}{M^0(z)} \ .$$

[3]See Chung and Olivares (2019, Section 4, and Appendix D) for alternative corrections. In particular, we also consider Holm (1979) and Westfall and Young (1993). As we argue in the main document, we present Bonferroni correction to make the comparisons with BGH's permutation test without prejudicing the latter.

Alternatively, the permutation test rejects the individual null hypothesis (3) if $K_{m,n,\hat{\delta}}(\hat{Z}^j)$ exceeds the $1-\alpha/\mathcal{J}$ quantile of the permutation distribution

$$\hat{R}_{m,n}^{K^j(\hat{\delta})}(t) = \frac{1}{M}\sum_{\pi\in\mathbf{G}_j} \mathbb{1}_{\{K^j_{m,n,\hat{\delta}}(\hat{Z}^j_\pi)\leq t\}} \ , \tag{7}$$

for all mutually exclusive $j\in\{1,\ldots,\mathcal{J}\}$.

1.4 Martingale-based Permutation Test

We begin by introducing more notation. Seee Chung and Olivares (2019, Section 3) for additional remarks, as well as the numerical and computational implementation of the martingale transformation. All the subsequent analysis is for a given subgroup $j\in\{1,\ldots,\mathcal{J}\}$.

Define the function $g(s) = (g_1(s), g_2(s)) = (s, f_0(F_0^{-1}(s)))'$ on $[0,1]$, and $\dot{g}(s) = (\dot{g}_1(s), \dot{g}_2(s))$ so that $\dot{g}$ is the derivative of g, where $F^{-1}(\tau) = \inf\{y : F(y)\geq\tau\}$, as usual. Therefore $\dot{g}(s) = (1, \dot{f}_0(F_0^{-1}(s))/f(F_0^{-1}(s)))$. Let $D[0,1]$ be the space of càdlàg functions on $[0,1]$, and denote by $\psi_g(h)(\cdot)$ the compensator of h, $\psi_g : D[0,1]\to D[0,1]$ given by

$$\psi_g(h)(t) = \int_0^t \left[\dot{g}(s)'C^{-1}(s)\int_s^1 \dot{g}(r)dh(r)\right]ds \tag{8}$$

where $C(s) = \int_s^1 \dot{g}(t)\dot{g}(t)'dt$. Lastly, consider the following equivalent transformation of (5) via the change of variable $y\mapsto F_0^{-1}(t)$ and work with

$$\begin{aligned} \upsilon^j_{m,n}(t,\hat{\delta}^j) &= \sqrt{\frac{m_j n_j}{N_j}}\left(\frac{1}{n_j}\sum_{i=1}^{n_j}\mathbb{1}_{\{\hat{Y}_i^{*j}(1)\leq F_0^{-1}(t)\}} - \frac{1}{m_j}\sum_{i=1}^{m_j}\mathbb{1}_{\{Y_i^j(1)\leq F_0^{-1}(t)\}}\right) \\ &= \sqrt{\frac{m_j n_j}{N_j}}\left(\hat{F}_1^{*j}(F_0^{-1}(t)) - \hat{F}_1^j(F_0^{-1}(t))\right) \\ &= V^j_{m,n}(F_0^{-1}(t),\hat{\delta}^j) \ . \end{aligned} \tag{9}$$

The two-sample martingale-transformed version of the two-sample Kolmogorov–Smirnov statistics is

$$\tilde{K}^j_{m,n,\hat{\delta}}(Z^j) = \sup_{0\leq t\leq 1}\left|\tilde{\upsilon}^j_{m,n}(t,\hat{\delta}^j)\right| \ . \tag{10}$$

We have that (10) converges in distribution to the supremum of a Brownian motion process $\mathbb{M}$, effectively nullifying the effect of the estimated nuisance parameter $\hat{\delta}^j$ (Chung and Olivares, 2019, Theorem 3). More importantly, the permutation distribution based on (10) behaves asymptotically like the true unconditional limiting distribution of $\tilde{K}^j_{m,n,\hat{\delta}}$, ensuring the asymptotic validity of the proposed permutation test (Chung and Olivares, 2019, Theorem 4).

2 Covariance Structures

As stated in Chung and Olivares (2019), let $\mathbb{G}$ and $\mathbb{B}$ be mean zero Brownian bridge and Brownian bridge with drift processes. More specifically, $\mathbb{G}$ has covariance function given by

$$\mathbb{E}\,\mathbb{G}(s)\mathbb{G}(t) = F_0(s \wedge t) - F_0(s)F_0(t)$$

As for the Brownian Bridge with drift, $\mathbb{B}$, let $\xi(\cdot)$ be a Gaussian process with mean 0 and covariance structure

$$\mathbb{C}(\xi(x), \xi(y)) = \sigma_0^2 f_0(x) f_0(y)$$

where $\sigma_0^2 = \sigma^2(F_0)$, and f_0 is the density of F_0. Then, the Brownian bridge with drift is given by $\mathbb{B}(\cdot) = \mathbb{G}(\cdot) + \xi(\cdot)$ with covariance structure

$$F_0(s \wedge t) - F_0(s)F_0(t) + \sigma_0^2 f_0(x) f_0(y) + 2\,\mathbb{C}(\mathbb{G}(x), \xi(y)) \ , \tag{11}$$

where the last summand is given by

$$\mathbb{C}(\mathbb{G}(x), \xi(y)) = f_0(y)F_0(x)\,(1 - F_0(x))\,\{\mathbb{E}(Y(0)|Y(0) \leq x) - \mathbb{E}(Y(0)|Y(0) > x)\} \ .$$

Throughout, let Φ and ϕ be respectively the cumulative probability distribution function and the probability density function of the standard normal distribution. Let $Y(0)$, be distributed as a lognormal distribution with (real) parameters μ and $\sigma > 0$, *i.e.* if Z is a standard normal variable then $Y(0)$ is given by

$$Y(0) = \mathrm{e}^{\mu + \sigma Z} \ .$$

Alternatively, $\ln\, Y(0) \ \sim \mathcal{N}(\mu, \sigma^2)$. The probability density function is given by

$$f_0(y) = \frac{1}{y\sigma}\phi\left(\frac{\ln\, y - \mu}{\sigma}\right) \ ,$$

and its cumulative distribution function by

$$F_0(y) = \Phi\left(\frac{\ln\, y - \mu}{\sigma}\right) \ .$$

The truncated moments are given by

$$\mathbb{E}(Y(0)|Y(0) \leq s) = \mathrm{e}^{\mu + \sigma^2/2}\left(\frac{\Phi\left(\frac{\ln s - \mu - \sigma^2}{\sigma}\right)}{\Phi\left(\frac{\ln s - \mu}{\sigma}\right)}\right)$$

$$\mathbb{E}(Y(0)|Y(0) > s) = \mathrm{e}^{\mu + \sigma^2/2}\left(\frac{\Phi\left(\frac{\mu + \sigma^2 - \ln s}{\sigma}\right)}{1 - \Phi\left(\frac{\ln s - \mu}{\sigma}\right)}\right) \ .$$

Therefore, the covariance $\mathbb{C}(\mathbb{G}(s), \xi(t))$ is given by

$$\begin{aligned}
\mathbb{C}(\mathbb{G}(s), \xi(t)) &= \left(\frac{\mathrm{e}^{\mu+\sigma^2/2}}{t\sigma}\right)\phi(t)\Phi\left(\frac{\ln s-\mu}{\sigma}\right)\left(1-\Phi\left(\frac{\ln s-\mu}{\sigma}\right)\right)\left(\frac{\Phi\left(\frac{\ln\ s-\mu-\sigma^2}{\sigma}\right)}{\Phi\left(\frac{\ln s-\mu}{\sigma}\right)} - \frac{\Phi\left(\frac{\mu+\sigma^2-\ln s}{\sigma}\right)}{1-\Phi\left(\frac{\ln s-\mu}{\sigma}\right)}\right) \\
&= \frac{\phi(t)\mathrm{e}^{\mu+\sigma^2/2}}{t\sigma}\left\{\Phi\left(\frac{\ln s-\mu-\sigma^2}{\sigma}\right) - \Phi\left(\frac{\ln s-\mu}{\sigma}\right)\right\} ,
\end{aligned}$$

where the last equality follows by the symmetry of the standard normal distribution.

3 Additional Empirical Results

Here we conduct the same empirical evaluation of Connecticut's Jobs First using the permutation test of Ding, Feller, and Miratrix (2015).

Table 1: Testing for Heterogeneity in the Treatment Effect by Subgroups, Time-varying mean treatment effects by subgroup with participation adjustment

		BGH's Permutation Test		FRT-CI		Martingale-based Permutation Test	
Subgroup	Number of Tests	Number of Reject at 10%	Number of Reject at 5%	Number of Reject at 10%	Number of Reject at 5%	Number of Reject at 10%	Number of Reject at 5%
Full Sample	7	4	4	3	3	7	7
Education	21	3	1	2	2	9	9
Age of youngest child	21	3	1	2	2	11	10
Marital status	21	2	1	2	1	14	14
Earnings level seventh Q pre-RA	21	2	1	1	1	17	16
Number of pre-RA Q with earnings	21	1	0	1	1	17	16
Welfare receipt seventh Q pre-RA	14	3	3	3	3	14	14
Education subgroups interacted with							
Age of youngest child	49	1	0	1	1	14	14
Marital status	35	3	3	2	2	18	17
Earnings level seventh Q pre-RA	63	1	0	0	0	15	14
Number of pre-RA Q with earnings	63	0	0	0	0	13	11
Welfare receipt seventh Q pre-RA	42	1	0	0	0	15	14
Age of youngest child interacted with							
Marital status	35	1	1	0	0	17	15
Earnings level seventh Q pre-RA	63	0	0	0	0	17	14
Number of pre-RA Q with earnings	49	1	1	0	0	14	12
Welfare receipt seventh Q pre-RA	42	1	0	1	0	14	13
Marital status subgroup interacted with							
Earnings level seventh Q pre-RA	63	2	1	0	0	14	11
Number of pre-RA Q with earnings	63	0	0	0	0	15	11
Welfare receipt seventh Q pre-RA	42	1	0	1	0	14	13
Earnings level seventh Q pre-RA subgroups interacted with							
Number of pre-RA Q with earnings	49	0	0	0	0	16	15
Welfare receipt seventh Q pre-RA	42	1	1	1	1	17	15
Number of quarters any earnings pre-RA subgroup interacted with							
Welfare receipt seventh Q pre-RA	42	0	0	0	0	14	13

All reported results account for multiple testing using Bonferroni adjustment. The martingale based permutation test was calculated based on 1000 permutations.

References

Chung, E. and Olivares, M. (2019). Permutation test for heterogeneous treatment effects with a nuisance parameter. *Working Paper*, pages 1–50. Latest version.

Ding, P., Feller, A., and Miratrix, L. (2015). Randomization inference for treatment effect variation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70.

Neyman, J. (1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. Originally published in 1923 and then translated by Dabrowska, Dorota M and Speed, TP. *Statistical Science*, pages 465–472.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.

Westfall, P. H. and Young, S. S. (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment*, volume 279. John Wiley & Sons.