The Benefits of Probability-Proportional-to-Size Sampling in Cluster-Randomized Experiments

Michael J. Higgins*

First Draft: 1/20/2014
This Draft: 9/19/2014 (08:45)

Abstract

In a cluster-randomized experiment, treatment is assigned to clusters of individual units of interest–households, classrooms, villages, etc.–instead of the units themselves. The number of clusters sampled and the number of units sampled within each cluster is typically restricted by a budget constraint. Previous analysis of cluster randomized experiments under the Neyman-Rubin potential outcomes model of response have assumed a simple random sample of clusters; estimators of the population average treatment effect (PATE) under this assumption are often either biased or not invariant to location shifts of potential outcomes. We demonstrate that, by sampling clusters with probability proportional to the number of units within a cluster, the PATE can be unbiasedly estimated using an estimator that satisfies unbiasedness and invariance to location shifts.

1 Introduction

Frequently in experiments, treatment is randomized across clusters, or groups, of units of interest instead of the units themselves—these are *cluster-randomized experiments* (CREs). Clusters of units are often formed *a priori* to the design of the experiment and without researcher intervention. Estimation of treatment effects is more precise when treatment is randomized across units [Cornfield, 1978], hence, logistical issues (rather than reducing variance of treatment effect estimates)

*Postdoctoral Research Associate, Program for Quantitative and Analytical Political Science, Princeton University. For the most recent draft, visit: http://www.princeton.edu/~mjh5/

motivate the randomization of treatment across clusters. Reasons for such randomization include addressing issues with the ethicality, legality, or feasibility of randomizing treatment across units, reducing risk of treatment contamination, and mimicking the implementation of a proposed program (e.g. a educational intervention that is implemented within classrooms) [Donner, 1998, Donner and Klar, 2004, Hayes and Moulton, 2009]. Common settings for cluster-randomized experiments include: testing an educational intervention that is implemented within classrooms [Raver et al., 2009]; evaluating efficacy of a health intervention that is implemented within clinics or medical practices [Bruce et al., 2004, King et al., 2007, Small et al., 2008, Imai et al., 2009]; measuring increases in compliance and turnout from mailers sent to households [Gerber and Green, 2000]; and identifying effects of interventions implemented within villages or other geographic regions [Wantchekon, 2003, Paluck, 2009, Beath et al., 2013].

To estimate and perform inference on the *population average treatment effect* (PATE), a CRE will require at least two stages of sampling: sampling clusters from a larger population of clusters (e.g. a sample of villages within a country) and sampling individual units from each of the sampled clusters—samples may be comprised of the entire sampling frame. After a sample of clusters is obtained, but before units are sampled within each cluster, treatment is allocated across sampled clusters. Researchers often improve the precision of treatment effect estimates by drawing a stratified sample of clusters and/or blocking sampled clusters before treatment assignment [Gail et al., 1996, Lewsey, 2004, Imai et al., 2009, Hayes and Moulton, 2009, Imbens, 2011, Hansen et al., 2014]. When researchers are interested in heterogeneous treatment effects across subpopulations of interest, within-cluster samples may also be stratified (for an example, see Kerry et al. [2005]).

When clusters are sampled using simple random sampling (SRS) or stratified random sampling (StRS), current estimators of the PATE have undesirable properties. The unbiased Horvitz-Thompson (HT) estimator [Horvitz and Thompson, 1952] is not invariant to location shifts of responses which inflates its variance. The location-invariant difference-in-means (DIM) estimator will be biased when treatment effects are correlated with cluster *sizes*—the number of units contained within each cluster [Middleton and Aronow, 2014]. Thus, this estimator is only unbiased in special cases such as under sharp null of no unit-level treatment effect [Hansen et al., 2014] or when clusters are blocked or stratified exactly on cluster sizes [Donner and Klar, 2004, Imai et al., 2009]. Moreover, when within-cluster samples are not drawn proportional to the cluster size, DIM may estimate a quantity different from the PATE. In fact, the only current estimator of the PATE that is both unbiased and location-invariant is the Des Raj estimator [Middleton and Aronow, 2014], which requires the introduction of an additional parameter; however, estimating this parameter will induce bias in the estimator.

We propose an adjustment in the *design* of the experiment—as opposed to adjusting weights of estimators after the experiment— for differences in cluster sizes: to sample clusters with *probability proportional to size* (PPS) [Hansen and Hurwitz, 1943, Cochran, 1977, Lohr, 1999]. We show that the Hansen-Hurwitz (HH) estimator of the PATE is both unbiased and location invariant when clusters are sampled using PPS sampling.

The paper is organized as follows: Section 2 introduces notation. Section 3 demonstrates that the HH estimator is both unbiased and location-invariant under PPS-without-replacement sampling of clusters and gives standard errors and estimates of standard errors for HH under this sampling scheme.

2 Notation, assumptions, and preliminaries

We consider an experiment with n units. Units are either assigned to treatment or control. Units are partitioned into $\#c$ clusters, numbered 1 through $\#c$. Let n_c denote the number of units with cluster c. Suppose units are ordered in some way within each cluster; let (k, c) denote the k^{th} unit in cluster c. We now introduce sampling and treatment assignment notation in the order in which they are performed in a CRE.

2.1 Sampling clusters

A total of $\#s$ clusters are sampled; we assume $\#s$ is fixed and chosen by the researcher. Let S_c denote a cluster sampling indicator; $S_c = 1$ if and only if cluster c contained in the sample.

$$S_c = \begin{cases} 1, & \text{cluster } c \text{ is sampled,} \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

By definition, $\sum_{c=1}^{\#c} S_c = \#s$.

2.2 Treatment assignment

Each of the $\#s$ sampled clusters is assigned to either treatment or control. Let T_{ct} denote a treatment indicator; $T_{ct} = 1$ if and only if cluster c receives treatment $t \in \{0, 1\}$.

$$T_{ct} = \begin{cases} 1, & \text{cluster } c \text{ receives treatment } t, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

We define $T_c = 0$ when $S_c = 0$. Let $\#T_t$ denote the number of clusters that receive treatment t.

We suppose that treatment assignment is *symmetric* across sampled clusters Miratrix et al. [2013]. That is, we suppose that, conditioned on the number of treated clusters $\#T_1$, each of the $\binom{\#s}{\#T_1}$ possible treatment assignments is equally likely. Symmetric treatment assignment implies that, for any treatment $t \in \{0, 1\}$ and distinct clusters c, c':

$$\mathbb{E}\left(T_{ct} \middle| \#T_t\right) = \frac{\#T_t}{\#s}, \tag{3}$$

$$\mathbb{E}\left(T_{ct}T_{c't} \middle| \#T_t\right) = \frac{\#T_t(\#T_t - 1)}{\#s(\#s - 1)}. \tag{4}$$

Complete randomization is a special case of symmetric treatment assignment. When the sample of clusters is stratified, symmetric treatment assignment also requires independence of treatment assignment across strata.

2.3 Within-cluster sampling

After treatment is assigned across clusters, a SRS of $\#s_c$ units is drawn within each sampled cluster c. This sample is drawn independently of treatment assignment and independently across clusters. We assume that these sample sizes are non-random and do not depend on the set of clusters sampled.

Let S_{kc} denote unit sampling indicator; $S_{kc} = 1$ if and only if the k^{th} unit in cluster c is sampled.

$$S_{kc} = \begin{cases} 1, & \text{unit } (k, c) \text{ is sampled,} \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

We define $S_{kc} = 0$ when $S_c = 0$. By definition, $\sum_{k=1}^{n_c} S_{kc} = \#s_c$.

2.4 Model of response: Neyman-Rubin Causal Model

Let y_{kct} denote the *potential outcome* of unit (k, c) given treatment t—the value of unit (k, c) we would have observed had that unit received treatment t. Note that y_{kct} is known if and only if that unit is sampled and receives treatment t; $S_c T_{ct} S_{kc} = 1$. Potential outcomes are assumed to be non-random. Let $\mathbf{y} = (y_{kct})_{k=1,\ c=1,\ t=0}^{n_c \quad \#c \quad 1}$ denote the vector of potential outcomes.

Let Y_{kc} denote the observed response of unit (k, c) had that unit been sampled. We assume responses follow the Neyman-Rubin Causal Model (NRCM) [Splawa-Neyman et al., 1990, Rubin, 1974, Holland, 1986]:

$$\begin{aligned} Y_{kc} &= y_{kc1} T_{c1} + y_{kc0} T_{c0} \\ &= y_{kc1} T_{c1} + y_{kc0}(1 - T_{c1}). \end{aligned} \tag{6}$$

Inherent in this model is the *stable-unit treatment value assumption* (SUTVA), which is often referred to as the *no-interference assumption*; the value of Y_{kc} only depends on the treatment assigned to cluster c and is not affected by the treatment assignment of any other cluster c'. Observe that this assumption only needs to hold across sampled clusters and does not need to hold for units within each cluster.

2.5 Parameter of interest and location invariance

Our quantity of interest is the *population average treatment effect* (PATE):

$$\delta = \delta(\mathbf{y}) \equiv \sum_{c=1}^{\#c} \sum_{k=1}^{n_c} \frac{y_{kc1} - y_{kc0}}{n} = \mu_1 - \mu_0, \tag{7}$$

where

$$\mu_t = \mu_t(\mathbf{y}) \equiv \sum_{c=1}^{\#c} \sum_{k=1}^{n_c} \frac{y_{kct}}{n} \tag{8}$$

denotes the population mean for treatment t. A transformation of potential outcomes $\mathbf{y} \rightarrow \mathbf{y}^*$ is *linear* if, for constants a, b:

$$y^*_{kct} = a + b y_{kct}, \quad \text{for all} \;\; k = 1, \ldots, n_c, \; c = 1, \ldots, \#c, \; t \in \{0, 1\}. \tag{9}$$

For simplicity, we may write this as $\mathbf{y}^* = a + b\mathbf{y}$. A *location transformation* or *shift* is a linear transformation in which $b = 1$.

Observe that the population mean is linear in potential outcomes:

$$\mu_t(\mathbf{a} + b\mathbf{y}) = a + b\mu_t(\mathbf{y}), \tag{10}$$

whereas the PATE is *location-invariant*—that is, the value does not change given a location shift of potential outcomes:

$$\delta(a + \mathbf{y}) = \delta(\mathbf{y}). \tag{11}$$

3 Estimation of PATE under PPS sampling

3.1 PPS sampling of clusters

Clusters are sampled with probability proportion to their *size* (PPS)—the number of units contained within the cluster. To be precise, we define a PPS sample with s draws as any sample in which the probability of any cluster c of being sampled is sn_c/n. While generally PPS samples can be drawn with replacement, we focus exclusively on PPS samples drawn without replacement (PPSWOR) where the number of unique clusters sampled are fixed. This allows researchers to have greater control in designing a CRE under a budget constraint. A PPSWOR sampling scheme requires each cluster to contain no more than n/s units.

Drawing a PPSWOR sample is a deceptively unintuitive task and quite a bit of work has been devoted to efficient and/or exact selection of PPSWOR samples [Hanurav, 1967, Vijayan, 1968, Sinha, 1973, Brewer and Hanif, 1982, Berger and Tillé, 2009]. Unlike SRS or sampling with replacement, PPSWOR sampling schemes are not uniquely defined solely by the property that the marginal probability of sampling a cluster is $n_c s/n$. Instead, for each pair of unique clusters c, c' a PPSWOR sampling scheme requires knowing the the joint probability $\pi_{cc'}$ of having both of these clusters included in the sample. To reduce variance in estimators, it is useful to choose a sampling scheme such that

$$\pi_{cc'} \geq P(S_c = 1)P(S_{c'} = 1) = n_c n_{c'} s^2/n^2 > 0. \tag{12}$$

[Sunter, 1986] provides a method of efficiently drawing an approximate PPSWOR sample satisfying (12).

3.2 Hansen-Hurwitz estimator

We define the Hansen-Hurwitz (HH) *sample mean* for treatment t as:

$$\hat{\mu}_{t,\text{HH}} = \hat{\mu}_t(\mathbf{y}) \equiv \sum_{c=1}^{\#c} S_c T_{ct} \frac{1}{\#T_t} \sum_{k=1}^{n_c} \frac{y_{kct} S_{kc}}{s_c}. \tag{13}$$

In words, this estimate is obtained by finding each cluster c that receives treatment t, computing the average response within each of these clusters, and then taking the average of these within-cluster averages. The HH estimator of the PATE is the difference of these sample means:

$$\hat{\delta}_{\text{HH}} = \hat{\mu}_{1,\text{HH}} - \hat{\mu}_{0,\text{HH}}. \tag{14}$$

Showing unbiasedness, deriving variances, and proving location invariance is straight forward under this sampling scheme. We now prove the following lemma:

Lemma 1 *Suppose that clusters are sampled according to PPSWOR sampling and suppose that treatment is symmetric across clusters. Then, for any treatment t,*

$$\mathbb{E}(\hat{\mu}_{t,HH}) = \mu_t \tag{15}$$

$$\begin{aligned} Var(\hat{\mu}_{t,HH}) &= \mathbb{E}\left(\frac{1}{\#T_t}\right)\left(\sigma^2_{t,bet} + \sigma^2_{t,with}\right) \\ &\quad + \mathbb{E}\left(\frac{1}{\#T_t} - 1\right)\left(\sum_{c=1}^{\#c}\sum_{c' \neq c} \frac{\pi_{cc'}\mu_{ct}\mu_{c't}}{s(s-1)} - \mu_t^2\right), \end{aligned} \tag{16}$$

$$cov(\hat{\mu}_{1,HH}, \hat{\mu}_{0,HH}) = \frac{1}{s(s-1)} \sum_{c=1}^{\#c}\sum_{c' \neq c} \pi_{cc'}\mu_{c1}\mu_{c'0} - \mu_1\mu_0. \tag{17}$$

Additionally, $\hat{\mu}_t$ is a linear function of potential outcomes:

$$\hat{\mu}_t(a + b\mathbf{y}) = a + b\hat{\mu}_t(\mathbf{y}) \tag{18}$$

Here, μ_{ct} and σ^2_{ct} denote the mean and the variance respectively of potential outcomes within cluster c under treatment t:

$$\mu_{ct} \equiv \sum_{k=1}^{n_c} \frac{y_{kct}}{n_c}, \tag{19}$$

$$\sigma^2_{ct} \equiv \sum_{k=1}^{n_c} \frac{(y_{kct} - \mu_{ct})^2}{n_c}; \tag{20}$$

the $\sigma^2_{t,\text{bet}}$ term denotes the weighted across-cluster variance for treatment t:

$$\sigma^2_{t,\text{bet}} \equiv \sum_{c=1}^{\#c} \frac{n_c}{n} \left(\mu_{ct} - \mu_t\right)^2; \tag{21}$$

and $\sigma^2_{t,\text{with}}$ denotes the weighted within-cluster variance for treatment t:

$$\sigma^2_{t,\text{with}} \equiv \sum_{c=1}^{\#c} \frac{n_c}{n} \left(\frac{n_c - s_c}{n_c - 1} \frac{\sigma^2_{ct}}{s_c} \right). \tag{22}$$

This lemma is proved in Appendices A and B.

Using this lemma, we now prove the following theorem:

Theorem 2 (Unbiasedness and location invariance) *Suppose that clusters are sampled accord-*

ing to PPSWOR sampling, and suppose that treatment is symmetric across clusters. Then:

$$\begin{aligned}
\mathbb{E}\left(\hat{\delta}_{HH}\right) &= \delta, \quad (23)\\
Var\left(\hat{\delta}_{HH}\right) &= \mathbb{E}\left(\frac{1}{\#T_1}\right)\left(\sigma^2_{1,bet}+\sigma^2_{1,with}\right)\\
&+\mathbb{E}\left(\frac{1}{\#T_1}-1\right)\left(\sum_{c=1}^{\#c}\sum_{c'\neq c}\frac{\pi_{cc'}\mu_{c1}\mu_{c'1}}{s(s-1)}-\mu_1^2\right)\\
&+\mathbb{E}\left(\frac{1}{\#T_0}\right)\left(\sigma^2_{0,bet}+\sigma^2_{0,with}\right)\\
&+\mathbb{E}\left(\frac{1}{\#T_0}-1\right)\left(\sum_{c=1}^{\#c}\sum_{c'\neq c}\frac{\pi_{cc'}\mu_{c0}\mu_{c'0}}{s(s-1)}-\mu_0^2\right)\\
&-\left(\frac{2}{s(s-1)}\sum_{c=1}^{\#c}\sum_{c'\neq c}\pi_{cc'}\mu_{c1}\mu_{c'0}-\mu_1\mu_0\right) \quad (24)
\end{aligned}$$

Moreover $\hat{\delta}_{HH}$ is invariant under location transformations of potential outcomes:

$$\hat{\delta}_{HH}(a+\mathbf{y}) = \hat{\delta}_{HH}(\mathbf{y}) \quad (25)$$

Proof: First we show location invariance. Observe from (18) that:

$$\begin{aligned}
\hat{\delta}_{\text{HH}}(a+\mathbf{y}) &= \hat{\mu}_{1,\text{HH}}(a+\mathbf{y})-\hat{\mu}_{0,\text{HH}}(a+\mathbf{y}) = a+\hat{\mu}_{1,\text{HH}}(\mathbf{y})-(a+\hat{\mu}_{0,\text{HH}}(\mathbf{y}))\\
&= \hat{\mu}_{1,\text{HH}}(\mathbf{y})-\hat{\mu}_{0,\text{HH}}(\mathbf{y}) = \hat{\delta}_{\text{HH}}(\mathbf{y}). \quad (26)
\end{aligned}$$

From (15) and linearity of expectations, we have:

$$\mathbb{E}\left(\hat{\delta}_{\text{HH}}\right) = \mathbb{E}\left(\hat{\mu}_{1,\text{HH}}-\hat{\mu}_{0,\text{HH}}\right) = \mathbb{E}\left(\hat{\mu}_{1,\text{HH}}\right)-\mathbb{E}\left(\hat{\mu}_{0,\text{HH}}\right) = \mu_1-\mu_0 = \delta. \quad (27)$$

The variance follows immediately from (16) and (17), and by observing that:

$$\mathrm{Var}\left(\hat{\delta}_{\mathrm{HH}}\right) = \mathrm{Var}\left(\hat{\mu}_{1,\mathrm{HH}} - \hat{\mu}_{0,\mathrm{HH}}\right) = \mathrm{Var}\left(\hat{\mu}_{1,\mathrm{HH}}\right) + \mathrm{Var}\left(\hat{\mu}_{0,\mathrm{HH}}\right) - 2\mathrm{cov}\left(\hat{\mu}_{1,\mathrm{HH}}, \hat{\mu}_{0,\mathrm{HH}}\right). \qquad (28)$$

■

References

Andrew Beath, Fotini Christia, and Ruben Enikolopov. Empowering women through development aid: Evidence from a field experiment in afghanistan. *American Political Science Review*, 107 (03):540–557, 2013.

Yves G Berger and Yves Tillé. Sampling with unequal probabilities. *Handbook of statistics*, 29: 39–54, 2009.

Kenneth RW Brewer and Muhammad Hanif. *Sampling with unequal probabilities (lecture notes in statistics)*. Springer, 1982.

Martha L Bruce, Thomas R Ten Have, Charles F Reynolds III, Ira I Katz, Herbert C Schulberg, Benoit H Mulsant, Gregory K Brown, Gail J McAvay, Jane L Pearson, and George S Alexopoulos. Reducing suicidal ideation and depressive symptoms in depressed older primary care patients: a randomized controlled trial. *Jama*, 291(9):1081–1091, 2004.

George Casella and Roger L Berger. *Statistical inference*, volume 70. Duxbury Press Belmont, CA, 1990.

W.G. Cochran. *Sampling techniques*. Wiley, New York, NY, 1977.

Jerome Cornfield. Randomization by group: A formal analysis. *American journal of epidemiology*, 108(2):100–102, 1978.

Allan Donner. Some aspects of the design and analysis of cluster randomization trials. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(1):95–113, 1998.

Allan Donner and Neil Klar. Pitfalls of and controversies in cluster randomization trials. *American Journal of Public Health*, 94(3):416–422, 2004.

Mitchell H Gail, Steven D Mark, Raymond J Carroll, Sylvan B Green, and David Pee. On design considerations and randomization-based inference for community intervention trials. *Statistics in medicine*, 15(11):1069–1092, 1996.

Alan S Gerber and Donald P Green. The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American Political Science Review*, pages 653–663, 2000.

Ben B Hansen, Paul R Rosenbaum, and Dylan S Small. Clustered treatment assignments and sensitivity to unmeasured biases in observational studies. *Journal of the American Statistical Association*, 109(505), 2014.

Morris H Hansen and William N Hurwitz. On the theory of sampling from finite populations. *The Annals of Mathematical Statistics*, 14(4):333–362, 1943.

TV Hanurav. Optimum utilization of auxiliary information: π ps sampling of two units from a stratum. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 374–391, 1967.

Richard Hayes and L Moulton. *Cluster randomised trials*. Chapman & Hall/CRC, 2009.

Paul W Holland. Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960, 1986.

Daniel G Horvitz and Donovan J Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952.

Kosuke Imai, Gary King, and Clayton Nall. The essential role of pair matching in cluster-randomized experiments, with application to the mexican universal health insurance evaluation. *Statistical Science*, 24(1):29–53, 2009.

Guido W. Imbens. Experimental design for unit and cluster randomized trials. Working Paper, 2011.

Sally M Kerry, Francesco P Cappuccio, Lynsey Emmett, Jacob Plange-Rhule, and John B Eastwood. Reducing selection bias in a cluster randomized trial in west african villages. *Clinical Trials*, 2(2):125–129, 2005.

Gary King, Emmanuela Gakidou, Nirmala Ravishankar, Ryan T Moore, Jason Lakin, Manett Vargas, Martha María Téllez-Rojo, Juan Eugenio Hernández Ávila, Mauricio Hernández Ávila, and Héctor Hernández Llamas. A politically robust experimental design for public policy evaluation, with application to the mexican universal health insurance program. *Journal of Policy Analysis and Management*, 26(3):479–506, 2007.

JD Lewsey. Comparing completely and stratified randomized designs in cluster randomized trials when the stratifying factor is cluster size: a simulation study. *Statistics in medicine*, 23(6): 897–905, 2004.

S.L. Lohr. *Sampling: Design and Analysis*. Duxbury Press, Pacific Grove, CA, 1999.

Joel A. Middleton and Peter M. Aronow. Unbiased estimation of the average treatment effect in cluster-randomized experiments. Submitted, 2014. URL `http://www.joelmiddleton.com/uploads/4/3/1/4/4314860/clusterrand-2011-12-24.pdf`.

Luke W. Miratrix, Jasjeet S. Sekhon, and Bin Yu. Adjusting treatment effect estimates by post-stratification in randomized experiments. *Journal of the Royal Statistical Society, Series B*, 75 (2):369–396, 2013.

Elizabeth Levy Paluck. Reducing intergroup prejudice and conflict using the media: a field experiment in rwanda. *Journal of personality and social psychology*, 96(3):574, 2009.

C Cybele Raver, Stephanie M Jones, Christine Li-Grining, Fuhua Zhai, Molly W Metzger, and Bonnie Solomon. Targeting children's behavior problems in preschool classrooms: a cluster-randomized controlled trial. *Journal of consulting and clinical psychology*, 77(2):302, 2009.

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

Bikas Kumar Sinha. On sampling schemes to realize preassigned sets of inclusion probabilities of first two orders. *Calcutta Statistical Association Bulletin*, 22:89–110, 1973.

Dylan S Small, Thomas R Ten Have, and Paul R Rosenbaum. Randomization inference in a group–randomized trial of treatments for depression: Covariate adjustment, noncompliance, and quantile effects. *Journal of the American Statistical Association*, 103(481):271–279, 2008.

Jerzy Splawa-Neyman, DM Dabrowska, and TP Speed. On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science*, 5(4):465–472, 1990.

Alan Sunter. Solutions to the problem of unequal probability sampling without replacement. *International statistical review*, 54(1):33–50, 1986.

K Vijayan. An exact πps sampling scheme-generalization of a method of hanurav. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 556–566, 1968.

Leonard Wantchekon. Clientelism and voting behavior: Evidence from a field experiment in benin. *World Politics*, 55(3):399–422, 2003.

A Location invariance

It follows that the estimator under this transformation of potential outcomes satisfies, for any constants a, b:

$$\begin{aligned}
\hat{\mu}_{t,\mathrm{HH}}(a + b\mathbf{y}) &= \sum_{c=1}^{\#c} \frac{S_c T_{ct}}{\#T_t} \sum_{k=1}^{n_c} \frac{(a + b y_{kct}) S_{kc}}{\# s_c} \\
&= \sum_{c=1}^{\#c} \frac{S_c T_{ct}}{\#T_t} \sum_{k=1}^{n_c} \frac{a S_{kc}}{s_c} + \sum_{c=1}^{\#c} \frac{S_c T_{ct}}{\#T_t} \sum_{k=1}^{n_c} \frac{b y_{kct} S_{kc}}{s_c} \\
&= a \sum_{c=1}^{\#c} \frac{S_c T_{ct}}{\#T_t} \sum_{k=1}^{n_c} \frac{S_{kc}}{s_c} + b \sum_{c=1}^{\#c} \frac{S_c T_{ct}}{\#T_t} \sum_{k=1}^{n_c} \frac{y_{kct} S_{kc}}{s_c} \\
&= a \sum_{c=1}^{\#c} \frac{S_c T_{ct}}{\#T_t} + b \hat{\mu}_{t,\mathrm{HH}}(\mathbf{y}) \\
&= a + b \hat{\mu}_{t,\mathrm{HH}}(\mathbf{y}). \qquad (29)
\end{aligned}$$

Hence, the HH sample mean is a linear function of potential outcomes.

B Expectation, variance, and covariance under PPSWOR

We begin this section by computing useful expectations under the assumptions in 2. Let $\mathbf{S} = (S_1, S_2, \ldots, S_b)$ denote a random set of cluster sampling indicator variables. For any distinct clusters c and c' and distinct treatments t and u, the following expectations hold under complete randomization of treatment to units, provided that $S_c = 1$ (and when applicable $S_{c'} = 1$):

$$
\begin{aligned}
\mathbb{E}\left(\frac{T_{ct}}{\#T_t}\middle|\mathbf{S}\right) &= \mathbb{E}\left[\mathbb{E}\left(\frac{T_{ct}}{\#T_t}\middle|\mathbf{S},\#T_t\right)\right] \\
&= \mathbb{E}\left(\frac{\frac{\#T_t}{s}}{\#T_t}\middle|\mathbf{S}\right) \\
&= \mathbb{E}\left(\frac{1}{s}\middle|\mathbf{S}\right) = \frac{1}{s}, && (30) \\
\mathbb{E}\left(\frac{T_{ct}}{(\#T_t)^2}\middle|\mathbf{S}\right) &= \mathbb{E}\left[\mathbb{E}\left(\frac{T_{ct}}{(\#T_t)^2}\middle|\mathbf{S},\#T_{ct}\right)\right] \\
&= \mathbb{E}\left(\frac{\frac{\#T_t}{s}}{(\#T_t)^2}\middle|\mathbf{S}\right) \\
&= \mathbb{E}\left(\frac{1}{s}\frac{1}{\#T_t}\middle|\mathbf{S}\right) = \frac{1}{s}\mathbb{E}\left(\frac{1}{\#T_t}\right), && (31) \\
\mathbb{E}\left(\frac{T_{ct}T_{c't}}{(\#T_t)^2}\middle|\mathbf{S}\right) &= \mathbb{E}\left[\mathbb{E}\left(\frac{T_{ct}T_{c't}}{(\#T_t)^2}\middle|\mathbf{S},\#T_t\right)\right] \\
&= \mathbb{E}\left(\frac{\frac{\#T_t}{s}\frac{\#T_t-1}{s-1}}{(\#T_t)^2}\middle|\mathbf{S}\right) \\
&= \mathbb{E}\left(\frac{(\#T_t)^2-\#T_t}{s(s-1)(\#T_t)^2}\middle|\mathbf{S}\right) \\
&= \frac{1}{s(s-1)}\mathbb{E}\left(1-\frac{1}{\#T_t}\right) \\
&= \frac{1}{s(s-1)} - \frac{1}{s(s-1)}\mathbb{E}\left(\frac{1}{\#T_t}\right), && (32) \\
\mathbb{E}\left(\frac{T_{c1}T_{c'0}}{\#T_t\#T_u}\middle|\mathbf{S}\right) &= \mathbb{E}\left[\mathbb{E}\left(\frac{T_{c1}T_{c'0}}{\#T_1\#T_0}\middle|\mathbf{S},\#T_1,\#T_0\right)\right] \\
&= \mathbb{E}\left(\frac{\frac{\#T_1}{s}\frac{\#T_0}{s-1}}{\#T_1\#T_0}\middle|\mathbf{S}\right) \\
&= \mathbb{E}\left(\frac{1}{s(s-1)}\right) = \frac{1}{s(s-1)}. && (33)
\end{aligned}
$$

Moreover, when units within a cluster are sampled using simple random sampling and within-cluster samples are drawn independently across clusters:

$$\mathbb{E}\left(S_{kc}\middle|\,\mathbf{S}\right) = \mathbb{E}\left(\frac{s_c}{n_c}\middle|\,\mathbf{S}\right) = \frac{s_c}{n_c} \tag{34}$$

$$\mathbb{E}\left(S_{kc}S_{k'c}\middle|\,\mathbf{S}\right) = \mathbb{E}\left(\frac{s_c}{n_c}\cdot\frac{s_c-1}{n_c-1}\middle|\,\mathbf{S}\right) = \frac{s_c(s_c-1)}{n_c(n_c-1)} \tag{35}$$

$$\mathbb{E}\left(S_{kc}S_{k^*c'}\middle|\,\mathbf{S}\right) = \mathbb{E}\left(\frac{s_c}{n_c}\cdot\frac{s_{c'}}{n_{c'}}\middle|\,\mathbf{S}\right) = \frac{s_c s_{c'}}{n_c n_{c'}} \tag{36}$$

When $S_c = 0$ (or, when applicable, $S_{c'} = 0$), these conditional expectations will equal zero. When the number of clusters sampled is fixed, $\#T_t$ is not dependent on $\mathbf{S}$.

From (31), (32), (33), (34), (35), and (36), and observing independence between the treatment assignment and the within-cluster sample, then when $c \neq c$, we obtain the following expectations:

$$\begin{aligned}
\mathbb{E}\left(\frac{S_cT_{ct}S_{kc}}{\#T_ts_c}\right) &= \mathbb{E}\left(\mathbb{E}\left(\frac{S_cT_{ct}S_{kc}}{\#T_ts_c}\middle|\,\mathbf{S}\right)\right) = \mathbb{E}\left(S_c\mathbb{E}\left(\frac{T_{ct}S_{kc}}{\#T_ts_c}\middle|\,\mathbf{S}\right)\right) \\
&= \mathbb{E}\left(S_c\,\mathbb{E}\left(\frac{S_{kc}}{s_c}\middle|\,\mathbf{S}\right)\mathbb{E}\left(\frac{T_{ct}}{\#T_t}\middle|\,\mathbf{S}\right)\right) \\
&= \mathbb{E}\left(S_c\,\mathbb{E}\left(\frac{\frac{s_c}{n_c}}{s_c}\middle|\,\mathbf{S}\right)\mathbb{E}\left(\mathbb{E}\left(\frac{T_{ct}}{\#T_t}\middle|\,\#T_t\right)\middle|\,\mathbf{S}\right)\right) \\
&= \mathbb{E}\left(S_c\mathbb{E}\left(\frac{1}{n_c}\middle|\,\mathbf{S}\right)\mathbb{E}\left(\frac{\frac{\#T_t}{s}}{\#T_t}\middle|\,\mathbf{S}\right)\right) \\
&= \frac{1}{n_c}\mathbb{E}\left(S_c\mathbb{E}\left(\frac{1}{s}\middle|\,\mathbf{S}\right)\right) = \frac{1}{n_cs}\mathbb{E}(S_c).
\end{aligned} \tag{37}$$

$$\begin{aligned}
\mathbb{E}\left(\frac{S_cT_{ct}S_{kc}}{(\#T_t)^2(st_c)^2}\right) &= \mathbb{E}\left(\mathbb{E}\left(\frac{S_cT_{ct}S_{kc}}{(\#T_t)^2(\#s_c)^2}\middle|\,\mathbf{S}\right)\right) = \mathbb{E}\left(S_c\mathbb{E}\left(\frac{T_{ct}S_{kc}}{(\#T_t)^2(\#s_c)^2}\middle|\,\mathbf{S}\right)\right) \\
&= \mathbb{E}\left(S_c\mathbb{E}\left(\frac{T_{ct}}{(\#T_t)^2}\middle|\,\mathbf{S}\right)\mathbb{E}\left(\frac{S_{kc}}{(\#s_c)^2}\middle|\,\mathbf{S}\right)\right) \\
&= \mathbb{E}\left(S_c\frac{1}{s}\mathbb{E}\left(\frac{1}{\#T_t}\right)\frac{\frac{s_c}{n_c}}{(\#s_c)^2}\right) = \mathbb{E}\left(\frac{1}{\#T_t}\right)\frac{1}{n_c\#s\#s_c}\mathbb{E}(S_c)
\end{aligned} \tag{38}$$

$$
\begin{aligned}
\mathbb{E}\left(\frac{S_c T_{ct} S_{kc} S_{k'c}}{(\#T_t)^2(\#s_c)^2}\right) &= \mathbb{E}\left(\mathbb{E}\left(\frac{S_c T_{ct} S_{kc} S_{k'c}}{(\#T_t)^2(\#s_c)^2}\middle|\mathbf{S}\right)\right) = \mathbb{E}\left(S_c \mathbb{E}\left(\frac{T_{ct} S_{kc} S_{k'c}}{(\#T_t)^2(\#s_c)^2}\middle|\mathbf{S}\right)\right) \\
&= \mathbb{E}\left(S_c \mathbb{E}\left(\frac{T_{ct}}{(\#T_t)^2}\middle|\mathbf{S}\right)\mathbb{E}\left(\frac{S_{kc}S_{k'c}}{(\#s_c)^2}\middle|\mathbf{S}\right)\right) \\
&= \mathbb{E}\left(S_c \frac{1}{\#s}\mathbb{E}\left(\frac{1}{\#T_t}\right)\frac{\frac{\#s_c(\#s_c-1)}{n_c(n_c-1)}}{(\#s_c)^2}\right) \\
&= \mathbb{E}\left(\frac{1}{\#T_t}\right)\frac{\#s_c-1}{n_c(n_c-1)\#s\#s_c}\mathbb{E}\left(S_c\right) \qquad (39)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left(\frac{S_c S_{c'} T_{ct} T_{c't} S_{kc} S_{k^*c'}}{(\#T_t)^2\#s_c\#s_{c'}}\right) &= \mathbb{E}\left(\mathbb{E}\left(\frac{S_c S_{c'} T_{ct} T_{c't} S_{kc} S_{k^*c'}}{(\#T_t)^2\#s_c\#s_{c'}}\middle|\mathbf{S}\right)\right) \\
&= \mathbb{E}\left(S_c S_{c'}\mathbb{E}\left(\frac{T_{ct} T_{c't} S_{kc} S_{k^*c'}}{(\#T_t)^2\#s_c\#s_{c'}}\middle|\mathbf{S}\right)\right) \\
&= \mathbb{E}\left(S_c S_{c'}\frac{1}{\#s(\#s-1)}\mathbb{E}\left(1-\frac{1}{\#T_t}\right)\frac{\frac{\#s_c\#s_{c'}}{n_c n_{c'}}}{\#s_c\#s_{c'}}\right) \\
&= \mathbb{E}\left(1-\frac{1}{\#T_t}\right)\frac{1}{\#s(\#s-1)n_c n_{c'}}\mathbb{E}\left(S_c S_{c'}\right) \qquad (40)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left(\frac{S_c S_{c'} T_{c1} T_{c'0} S_{kc} S_{k^*c'}}{\#T_1\#T_0\#s_c\#s_{c'}}\right) &= \mathbb{E}\left(\mathbb{E}\left(\frac{T_{c1} T_{c'0} S_{kc} S_{k^*c'}}{\#T_1\#T_0\#s_c\#s_{c'}}\middle|\mathbf{S}\right)\right) \\
&= \mathbb{E}\left(S_c S_{c'}\mathbb{E}\left(\frac{T_{c1} T_{c'0} S_{kc} S_{k^*c'}}{\#T_1\#T_0\#s_c\#s_{c'}}\middle|\mathbf{S}\right)\right) \\
&= \mathbb{E}\left(S_c S_{c'}\frac{1}{\#s(\#s-1)}\frac{\frac{\#s_c\#s_{c'}}{n_c n_{c'}}}{\#s_c\#s_{c'}}\right) \\
&= \frac{1}{\#s(\#s-1)n_c n_{c'}}\mathbb{E}\left(S_c S_{c'}\right) \qquad (41)
\end{aligned}
$$

Observe that, under PPSWOR sampling, $E(S_c) = n_c\#s/n$. Define $\pi_{cc'} \equiv E(S_c S_{c'}) = P(S_c = 1, S_{c'} = 1)$ as the probability of sampling both cluster c and c'. Hence, (37), (38), (39), (40), and (41) simplify to:

$$\begin{aligned}\mathbb{E}\left(\frac{S_cT_{ct}S_{kc}}{\#T_t\#s_c}\right) &= \frac{1}{n_c}\mathbb{E}\left(S_c\mathbb{E}\left(\frac{1}{\#s}\middle|\mathbf{S}\right)\right) = \frac{1}{n_c\#s}\mathbb{E}(S_c)\\ &= \frac{1}{n_c\#s}\frac{n_c\#s}{n} = \frac{1}{n}\end{aligned} \tag{42}$$

$$\begin{aligned}\mathbb{E}\left(\frac{S_cT_{ct}S_{kc}}{(\#T_t)^2(\#s_c)^2}\right) &= \mathbb{E}\left(\frac{1}{\#T_t}\right)\frac{1}{n_c\#s\#s_c}\mathbb{E}(S_c)\\ &= \mathbb{E}\left(\frac{1}{\#T_t}\right)\frac{1}{n_c\#s\#s_c}\frac{n_c\#s}{n} = \mathbb{E}\left(\frac{1}{\#T_t}\right)\frac{1}{n\#s_c}\end{aligned} \tag{43}$$

$$\begin{aligned}\mathbb{E}\left(\frac{S_cT_{ct}S_{kc}S_{k'c}}{(\#T_t)^2(\#s_c)^2}\right) &= \mathbb{E}\left(\frac{1}{\#T_t}\right)\frac{\#s_c-1}{n_c(n_c-1)\#s\#s_c}\mathbb{E}\left(S_c\right)\\ &= \mathbb{E}\left(\frac{1}{\#T_t}\right)\frac{\#s_c-1}{n_c(n_c-1)\#s\#s_c}\frac{n_c\#s}{n} = \frac{\#s_c-1}{n(n_c-1)\#s_c}\end{aligned} \tag{44}$$

$$\begin{aligned}\mathbb{E}\left(\frac{S_cS_{c'}T_{ct}T_{c't}S_{kc}S_{k^*c'}}{(\#T_t)^2\#s_c\#s_{c'}}\right) &= \mathbb{E}\left(1-\frac{1}{\#T_t}\right)\frac{1}{\#s(\#s-1)n_cn_{c'}}\mathbb{E}\left(S_cS_{c'}\right)\\ &= \mathbb{E}\left(1-\frac{1}{\#T_t}\right)\frac{\pi_{cc'}}{\#s(\#s-1)n_cn_{c'}}\end{aligned} \tag{45}$$

$$\begin{aligned}\mathbb{E}\left(\frac{S_cS_{c'}T_{c1}T_{c'0}S_{kc}S_{k^*c'}}{\#T_1\#T_0\#s_c\#s_{c'}}\right) &= \frac{1}{\#s(\#s-1)n_cn_{c'}}\mathbb{E}\left(S_cS_{c'}\right)\\ &= \frac{\pi_{cc'}}{\#s(\#s-1)n_cn_{c'}}\end{aligned} \tag{46}$$

We apply (42) to show unbiasedness of the sample mean:

$$\begin{aligned}\mathbb{E}(\hat{\mu}_{t,\mathrm{HH}}) &= \mathbb{E}\left(\sum_{c=1}^{\#c}\frac{S_cT_{ct}}{\#T_t}\sum_{k=1}^{n_c}\frac{y_{kct}S_{kc}}{\#s_c}\right)\\ &= \sum_{c=1}^{\#c}\sum_{k=1}^{n_c}y_{kct}\mathbb{E}\left(\frac{\#S_cT_{ct}S_{kc}}{\#T_ts_c}\right) = \sum_{c=1}^{\#c}\sum_{k=1}^{n_c}\frac{y_{kct}}{n} = \mu_t.\end{aligned} \tag{47}$$

We derive the variance $\text{Var}(\hat{\mu}_t)$ through the relationship:

$$\text{Var}(\hat{\mu}_{t,\text{HH}}) = \mathbb{E}\left(\hat{\mu}^2_{t,\text{HH}}\right) - \left(\mathbb{E}\left(\hat{\mu}_{t,\text{HH}}\right)\right)^2. \tag{48}$$

We have already shown that $\mathbb{E}(\hat{\mu}_{t,\text{HH}}) = \mu_t$. We now derive $\mathbb{E}(\hat{\mu}^2_{t,\text{HH}})$.

Note that:

$$\begin{aligned}
\hat{\mu}^2_{t,\text{HT}} &= \left(\sum_{c=1}^{\#c} \frac{S_c T_{ct}}{\#T_t} \sum_{k=1}^{n_c} \frac{y_{kct} S_{kc}}{\#s_c}\right)^2 \\
&= \sum_{c=1}^{\#c} \left(\frac{S_c T_{ct}}{\#T_t} \sum_{k=1}^{n_c} \frac{y_{kct} S_{kc}}{\#s_c}\right)^2 + \sum_{c=1}^{\#c} \sum_{c' \neq c} \left(\frac{S_c T_{ct}}{\#T_t} \sum_{k=1}^{n_c} \frac{y_{kct} S_{kc}}{\#s_c}\right) \left(\frac{S_{c'} T_{c't}}{\#T_t} \sum_{k^*=1}^{n_{c'}} \frac{y_{k^*c't} S_{k^*c'}}{\#s_{c'}}\right) \\
&= \sum_{c=1}^{\#c} \frac{S_c^2 T_{ct}^2}{(\#T_t)^2} \sum_{k=1}^{n_c} \left(\frac{y_{kct} S_{kc}}{\#s_c}\right)^2 + \sum_{c=1}^{\#c} \frac{S_c^2 T_{ct}^2}{(\#T_t)^2} \sum_{k=1}^{n_c} \sum_{k' \neq k} \left(\frac{y_{kct} S_{kc}}{\#s_c}\right) \left(\frac{y_{k'ct} S_{k'c}}{\#s_c}\right) \\
&\quad + \sum_{c=1}^{\#c} \sum_{c' \neq c} \left(\frac{S_c T_{ct}}{\#T_t} \sum_{k=1}^{n_c} \frac{y_{kct} S_{kc}}{\#s_c}\right) \left(\frac{S_{c'} T_{c't}}{\#T_t} \sum_{k^*=1}^{n_{c'}} \frac{y_{k^*c't} S_{k^*c'}}{\#s_{c'}}\right)
\end{aligned} \tag{49}$$

We continue:

$$
\begin{aligned}
& \sum_{c=1}^{\#c} \frac{S_c^2 T_{ct}^2}{(\#T_t)^2} \sum_{k=1}^{n_c} \left(\frac{y_{kct} S_{kc}}{\#s_c}\right)^2 + \sum_{c=1}^{\#c} \frac{S_c^2 T_{ct}^2}{(\#T_t)^2} \sum_{k=1}^{n_c} \sum_{k' \neq k} \left(\frac{y_{kct} S_{kc}}{\#s_c}\right) \left(\frac{y_{k'ct} S_{k'c}}{\#s_c}\right) \\
& + \sum_{c=1}^{\#c} \sum_{c' \neq c} \left(\frac{S_c T_{ct}}{\#T_t} \sum_{k=1}^{n_c} \frac{y_{kct} S_{kc}}{\#s_c}\right) \left(\frac{S_{c'} T_{c't}}{\#T_t} \sum_{k^*=1}^{n_{c'}} \frac{y_{k^*c't} S_{k^*c'}}{\#s_{c'}}\right) \\
= & \sum_{c=1}^{\#c} \frac{S_c^2 T_{ct}^2}{(\#T_t)^2} \sum_{k=1}^{n_c} \frac{y_{kct}^2 S_{kc}^2}{(\#s_c)^2} + \sum_{c=1}^{\#c} \frac{S_c^2 T_{ct}^2}{(\#T_t)^2} \sum_{k=1}^{n_c} \sum_{k' \neq k} \frac{y_{kct} y_{k'ct} S_{kc} S_{k'c}}{(\#s_c)^2} \\
& + \sum_{c=1}^{\#c} \sum_{c' \neq c} \frac{S_c S_{c'} T_{ct} T_{c't}}{(\#T_t)^2} \sum_{k=1}^{n_c} \sum_{k^*=1}^{n_{c'}} \frac{y_{kct} y_{k^*c't} S_{kc} S_{k^*c'}}{\#s_c \#s_{c'}} \\
= & \sum_{c=1}^{\#c} \sum_{k=1}^{n_c} \frac{y_{kct}^2 S_c^2 T_{ct}^2 S_{kc}^2}{(\#T_t)^2 (\#s_c)^2} + \sum_{c=1}^{\#c} \sum_{k=1}^{n_c} \sum_{k' \neq k} \frac{y_{kct} y_{k'ct} S_c^2 T_{ct}^2 S_{kc} S_{k'c}}{(\#T_t)^2 (\#s_c)^2} \\
& + \sum_{c=1}^{\#c} \sum_{c' \neq c} \sum_{k=1}^{n_c} \sum_{k^*=1}^{n_{c'}} \frac{y_{kct} y_{k^*c't} S_c S_{c'} T_{ct} T_{c't} S_{kc} S_{k^*c'}}{(\#T_t)^2 \#s_c \#s_{c'}} \\
= & \sum_{c=1}^{\#c} \sum_{k=1}^{n_c} \frac{y_{kct}^2 S_c T_{ct} S_{kc}}{(\#T_t)^2 (\#s_c)^2} + \sum_{c=1}^{\#c} \sum_{k=1}^{n_c} \sum_{k' \neq k} \frac{y_{kct} y_{k'ct} S_c T_{ct} S_{kc} S_{k'c}}{(\#T_t)^2 (\#s_c)^2} \\
& + \sum_{c=1}^{\#c} \sum_{c' \neq c} \sum_{k=1}^{n_c} \sum_{k^*=1}^{n_{c'}} \frac{y_{kct} y_{k^*c't} S_c S_{c'} T_{ct} T_{c't} S_{kc} S_{k^*c'}}{(\#T_t)^2 \#s_c \#s_{c'}}
\end{aligned} \tag{50}
$$

From (50), and applying (38), (39), and (40), we obtain the expectation:

$$
\begin{aligned}
\mathbb{E}\left(\hat{\mu}_{t,\text{HT}}^2\right) &= \mathbb{E}\left(\sum_{c=1}^{\#c}\sum_{k=1}^{n_c}\frac{y_{kct}^2 S_c T_{ct} S_{kc}}{(\#T_t)^2(\#s_c)^2} + \sum_{c=1}^{\#c}\sum_{k=1}^{n_c}\sum_{k'\neq k}\frac{y_{kct}y_{k'ct}S_c T_{ct} S_{kc} S_{k'c}}{(\#T_t)^2(\#s_c)^2}\right) \\
&\quad + \mathbb{E}\left(\sum_{c=1}^{\#c}\sum_{c'\neq c}\sum_{k=1}^{n_c}\sum_{k^*=1}^{n_{c'}}\frac{y_{kct}y_{k^*c't}S_c S_{c'} T_{ct} T_{c't} S_{kc} S_{k^*c'}}{(\#T_t)^2\#s_c\#s_{c'}}\right) \\
&= \sum_{c=1}^{\#c}\sum_{k=1}^{n_c} y_{kct}^2\mathbb{E}\left(\frac{S_c T_{ct} S_{kc}}{(\#T_t)^2(\#s_c)^2}\right) + \sum_{c=1}^{\#c}\sum_{k=1}^{n_c}\sum_{k'\neq k} y_{kct}y_{k'ct}\mathbb{E}\left(\frac{S_c T_{ct} S_{kc} S_{k'c}}{(\#T_t)^2(\#s_c)^2}\right) \\
&\quad + \sum_{c=1}^{\#c}\sum_{c'\neq c}\sum_{k=1}^{n_c}\sum_{k^*=1}^{n_{c'}} y_{kct}y_{k^*c't}\mathbb{E}\left(\frac{S_c S_{c'} T_{ct} T_{c't} S_{kc} S_{k^*c'}}{(\#T_t)^2\#s_c\#s_{c'}}\right) \\
&= \sum_{c=1}^{\#c}\sum_{k=1}^{n_c} y_{kct}^2\mathbb{E}\left(\frac{1}{\#T_t}\right)\frac{1}{n\#s_c} + \sum_{c=1}^{\#c}\sum_{k=1}^{n_c}\sum_{k'\neq k} y_{kct}y_{k'ct}\mathbb{E}\left(\frac{1}{\#T_t}\right)\frac{\#s_c-1}{n\#s_c(n_c-1)} \\
&\quad + \sum_{c=1}^{\#c}\sum_{c'\neq c}\sum_{k=1}^{n_c}\sum_{k^*=1}^{n_{c'}} y_{kct}y_{k^*c't}\mathbb{E}\left(1-\frac{1}{\#T_t}\right)\frac{\pi_{cc'}}{s(s-1)n_c n_{c'}} \\
&= \frac{1}{n}\mathbb{E}\left(\frac{1}{\#T_t}\right)\sum_{c=1}^{\#c}\sum_{k=1}^{n_c} y_{kct}^2\frac{1}{\#s_c} + \frac{1}{n}\mathbb{E}\left(\frac{1}{\#T_t}\right)\sum_{c=1}^{\#c}\frac{1}{n_c-1}\sum_{k=1}^{n_c}\sum_{k'\neq k} y_{kct}y_{k'ct}\left(1-\frac{1}{\#s_c}\right) \\
&\quad + \frac{1}{s(s-1)}\mathbb{E}\left(1-\frac{1}{\#T_t}\right)\sum_{c=1}^{\#c}\sum_{c'\neq c}\frac{\pi_{cc'}}{n_c n_{c'}}\sum_{k=1}^{n_c}\sum_{k^*=1}^{n_{c'}} y_{kct}y_{k^*c't}
\end{aligned} \tag{51}
$$

We simplify further:

$$
\begin{aligned}
& \frac{1}{n}\mathbb{E}\left(\frac{1}{\#T_t}\right)\sum_{c=1}^{\#c}\sum_{k=1}^{n_c} y_{kct}^2\frac{1}{\#s_c} + \frac{1}{n}\mathbb{E}\left(\frac{1}{\#T_t}\right)\sum_{c=1}^{\#c}\frac{1}{n_c-1}\sum_{k=1}^{n_c}\sum_{k'\neq k} y_{kct}y_{k'ct}\left(1-\frac{1}{\#s_c}\right) \\
& +\frac{1}{s(s-1)}\mathbb{E}\left(1-\frac{1}{\#T_t}\right)\sum_{c=1}^{\#c}\sum_{c'\neq c}\frac{\pi_{cc'}}{n_c n_{c'}}\sum_{k=1}^{n_c}\sum_{k^*=1}^{n_{c'}} y_{kct}y_{k^*c't} \\
= & \frac{1}{n}\mathbb{E}\left(\frac{1}{\#T_t}\right)\left(\sum_{c=1}^{\#c}\sum_{k=1}^{n_c} y_{kct}^2\frac{1}{\#s_c} + \sum_{c=1}^{\#c}\frac{1}{n_c-1}\sum_{k=1}^{n_c}\sum_{k'\neq k} y_{kct}y_{k'ct}\left(1-\frac{1}{\#s_c}\right)\right) \\
& -\frac{1}{s(s-1)}\mathbb{E}\left(1-\frac{1}{\#T_t}\right)\sum_{c=1}^{\#c}\sum_{c'\neq c}\pi_{cc'}\sum_{k=1}^{n_c}\frac{y_{kct}}{n_c}\sum_{k^*=1}^{n_{c'}}\frac{y_{k^*c't}}{n_c} \\
= & \frac{1}{n}\mathbb{E}\left(\frac{1}{\#T_t}\right)\left(\sum_{c=1}^{\#c}\sum_{k=1}^{n_c} y_{kct}^2\frac{1}{\#s_c} + \sum_{c=1}^{\#c}\frac{1}{n_c-1}\sum_{k=1}^{n_c}\sum_{k'\neq k} y_{kct}y_{k'ct}\left(1-\frac{1}{\#s_c}\right)\right) \\
& -\frac{1}{s(s-1)}\mathbb{E}\left(1-\frac{1}{\#T_t}\right)\sum_{c=1}^{\#c}\sum_{c'\neq c}\pi_{cc'}\mu_{ct}\mu_{c't}
\end{aligned} \tag{52}
$$

We focus on the term in parenthesis.

$$
\begin{aligned}
& \sum_{c=1}^{\#c}\sum_{k=1}^{n_c} y_{kct}^2\frac{1}{\#s_c} + \sum_{c=1}^{\#c}\frac{1}{n_c-1}\sum_{k=1}^{n_c}\sum_{k'\neq k} y_{kct}y_{k'ct}\left(1-\frac{1}{\#s_c}\right) \\
= & \sum_{c=1}^{\#c}\sum_{k=1}^{n_c} y_{kct}^2\frac{1}{\#s_c} + \sum_{c=1}^{\#c}\frac{1}{n_c-1}\left(1-\frac{1}{\#s_c}\right)\left(\left(\sum_{k=1}^{n_c} y_{kct}\right)^2 - \sum_{k=1}^{n_c} y_{kct}^2\right) \\
= & \sum_{c=1}^{\#c}\left(\frac{1}{\#s_c}\sum_{k=1}^{n_c} y_{kct}^2 - \frac{1}{n_c-1}\left(1-\frac{1}{\#s_c}\right)\sum_{k=1}^{n_c} y_{kct}^2\right) \\
& +\sum_{c=1}^{\#c}\frac{1}{n_c-1}\left(1-\frac{1}{\#s_c}\right)\left(\sum_{k=1}^{n_c} y_{kct}\right)^2 \\
= & \sum_{c=1}^{\#c}\left(\frac{n_c-1}{(n_c-1)\#s_c}\sum_{k=1}^{n_c} y_{kct}^2 - \frac{\#s_c-1}{(n_c-1)\#s_c}\sum_{k=1}^{n_c} y_{kct}^2 + \frac{\#s_c-1}{(n_c-1)\#s_c}\left(\sum_{k=1}^{n_c} y_{kct}\right)^2\right) \\
= & \sum_{c=1}^{\#c}\left(\frac{n_c-\#s_c}{(n_c-1)\#s_c}\sum_{k=1}^{n_c} y_{kct}^2 + \frac{\#s_c-1}{(n_c-1)\#s_c}\left(\sum_{k=1}^{n_c} y_{kct}\right)^2\right)
\end{aligned} \tag{53}
$$

Focusing on the term in parenthesis in the previous expression:

$$
\begin{aligned}
& \frac{n_c - \#s_c}{(n_c-1)\#s_c} \sum_{k=1}^{n_c} y_{kct}^2 + \frac{\#s_c - 1}{(n_c-1)\#s_c} \left(\sum_{k=1}^{n_c} y_{kct} \right)^2 \\
= \; & \frac{n_c(n_c - \#s_c)}{(n_c-1)\#s_c} \sum_{k=1}^{n_c} \frac{y_{kct}^2}{n_c} + \frac{n_c^2(\#s_c - 1)}{(n_c-1)\#s_c} \left(\sum_{k=1}^{n_c} \frac{y_{kct}}{n_c} \right)^2 \\
& + n_c \left(\sum_{k=1}^{n_c} \frac{y_{kct}}{n_c} \right)^2 - n_c \left(\sum_{k=1}^{n_c} \frac{y_{kct}}{n_c} \right)^2 \\
= \; & n_c \left(\sum_{k=1}^{n_c} \frac{y_{kct}}{n_c} \right)^2 + \frac{n_c(n_c - \#s_c)}{(n_c-1)\#s_c} \sum_{k=1}^{n_c} \frac{y_{kct}^2}{n_c} + \frac{n_c^2(\#s_c - 1) - n_c(n_c-1)\#s_c}{(n_c-1)\#s_c} \left(\sum_{k=1}^{n_c} \frac{y_{kct}}{n_c} \right)^2 \\
= \; & n_c \left(\sum_{k=1}^{n_c} \frac{y_{kct}}{n_c} \right)^2 + \frac{n_c(n_c - \#s_c)}{(n_c-1)\#s_c} \sum_{k=1}^{n_c} \frac{y_{kct}^2}{n_c} - \frac{n_c^2\#s_c - n_c\#s_c - n_c^2\#s_c + n_c^2}{(n_c-1)\#s_c} \left(\sum_{k=1}^{n_c} \frac{y_{kct}}{n_c} \right)^2 \\
= \; & n_c \left(\sum_{k=1}^{n_c} \frac{y_{kct}}{n_c} \right)^2 + \frac{n_c(n_c - \#s_c)}{(n_c-1)\#s_c} \sum_{k=1}^{n_c} \frac{y_{kct}^2}{n_c} - \frac{n_c(n_c - \#s_c)}{(n_c-1)\#s_c} \left(\sum_{k=1}^{n_c} \frac{y_{kct}}{n_c} \right)^2 \\
= \; & n_c \left(\sum_{k=1}^{n_c} \frac{y_{kct}}{n_c} \right)^2 + \frac{n_c(n_c - \#s_c)}{(n_c-1)\#s_c} \sigma_{ct}^2 = n_c \left(\mu_{ct}^2 + \frac{n_c - \#s_c}{n_c - 1} \frac{\sigma_{ct}^2}{\#s_c} \right) \qquad (54)
\end{aligned}
$$

Observe that this is a square of the cluster mean plus the variance of the cluster sample mean from a simple random sample of units from the cluster, multiplied by the number of units within the cluster. From (53) and (54), we find:

$$
\begin{aligned}
& \sum_{c=1}^{\#c} \left(\frac{n_c - \#s_c}{(n_c-1)\#s_c} \sum_{k=1}^{n_c} y_{kct}^2 + \frac{\#s_c - 1}{(n_c-1)\#s_c} \left(\sum_{k=1}^{n_c} y_{kct} \right)^2 \right) = \\
= \; & \sum_{c=1}^{\#c} n_c \left(\mu_{ct}^2 + \frac{n_c - \#s_c}{n_c - 1} \frac{\sigma_{ct}^2}{\#s_c} \right) \qquad (55)
\end{aligned}
$$

Hence, from (52)

$$\begin{aligned}
& \frac{1}{n}\mathbb{E}\left(\frac{1}{\#T_t}\right)\left(\sum_{c=1}^{\#c}\sum_{k=1}^{n_c} y_{kct}^2 \frac{1}{\#s_c} + \sum_{c=1}^{\#c}\frac{1}{n_c-1}\sum_{k=1}^{n_c}\sum_{k'\neq k} y_{kct}y_{k'ct}\left(1-\frac{1}{\#s_c}\right)\right) \\
& -\frac{1}{s(s-1)}\mathbb{E}\left(1-\frac{1}{\#T_t}\right)\sum_{c=1}^{\#c}\sum_{c'\neq c}\pi_{cc'}\mu_c\mu_{c'} \\
= \; & \frac{1}{n}\mathbb{E}\left(\frac{1}{\#T_t}\right)\left(\sum_{c=1}^{\#c} n_c\left(\mu_{ct}^2 + \frac{n_c-\#s_c}{n_c-1}\frac{\sigma_{\mathrm{ct}}^2}{\#s_c}\right)\right) \\
& -\frac{1}{s(s-1)}\mathbb{E}\left(1-\frac{1}{\#T_t}\right)\sum_{c=1}^{\#c}\sum_{c'\neq c}\pi_{cc'}\mu_{ct}\mu_{c't} \qquad (56)
\end{aligned}$$

And so:

$$\begin{aligned}
\mathrm{Var}(\hat{\mu}_{t,\mathrm{HT}}) &= \mathbb{E}\left(\hat{\mu}_{t,\mathrm{HT}}^2\right) - \left(\mathbb{E}\left(\hat{\mu}_{t,\mathrm{HT}}\right)\right)^2 = \mathbb{E}\left(\hat{\mu}_{t,\mathrm{HT}}^2\right) - \mu_t^2 \\
&= \frac{1}{n}\mathbb{E}\left(\frac{1}{\#T_t}\right)\left(\sum_{c=1}^{\#c} n_c\left(\mu_{ct}^2 + \frac{n_c-\#s_c}{n_c-1}\frac{\sigma_{\mathrm{ct}}^2}{\#s_c}\right)\right) - \mu_{t,\mathrm{HT}}^2 \\
& \quad -\frac{1}{s(s-1)}\mathbb{E}\left(1-\frac{1}{\#T_t}\right)\sum_{c=1}^{\#c}\sum_{c'\neq c}\pi_{cc'}\mu_{ct}\mu_{c't} \qquad (57)
\end{aligned}$$

We can simplify this further:

$$\begin{aligned}
\mathrm{Var}(\hat{\mu}_{t,\mathrm{HT}}) &= \frac{1}{n}\mathbb{E}\left(\frac{1}{\#T_t}\right)\left(\sum_{c=1}^{\#c} n_c\left(\mu_{ct}^2+\frac{n_c-\#s_c}{n_c-1}\frac{\sigma_{ct}^2}{\#s_c}\right)\right)-\mu_t^2 \\
&\quad -\frac{1}{s(s-1)}\mathbb{E}\left(1-\frac{1}{\#T_t}\right)\sum_{c=1}^{\#c}\sum_{c'\neq c}\pi_{cc'}\mu_{ct}\mu_{c't} \\
&= \mathbb{E}\left(\frac{1}{\#T_t}\right)\left(\sum_{c=1}^{\#c}\frac{n_c}{n}\left(\mu_{ct}^2+\frac{n_c-\#s_c}{n_c-1}\frac{\sigma_{ct}^2}{\#s_c}\right)\right)-\mathbb{E}\left(\frac{1}{\#T_t}\right)\mu_t^2 \\
&\quad -\mathbb{E}\left(1-\frac{1}{\#T_t}\right)\sum_{c=1}^{\#c}\sum_{c'\neq c}\frac{\pi_{cc'}\mu_{ct}\mu_{c't}}{s(s-1)}-\mathbb{E}\left(1-\frac{1}{\#T_t}\right)\mu_t^2 \\
&= \mathbb{E}\left(\frac{1}{\#T_t}\right)\left(\sum_{c=1}^{\#c}\frac{n_c}{n}\left(\mu_{ct}^2+\frac{n_c-\#s_c}{n_c-1}\frac{\sigma_{ct}^2}{\#s_c}\right)-\mu_t^2\right) \\
&\quad +\mathbb{E}\left(\frac{1}{\#T_t}-1\right)\left(\sum_{c=1}^{\#c}\sum_{c'\neq c}\frac{\pi_{cc'}\mu_{ct}\mu_{c't}}{s(s-1)}+\mu_t^2\right) \qquad (58)
\end{aligned}$$

Now,

$$\begin{aligned}
\frac{1}{n}\sum_{c=1}^{\#c} n_c\mu_{ct}^2-\mu_t^2 &= \frac{1}{n}\sum_{c=1}^{\#c} n_c\mu_{ct}^2-2\mu_t\mu_t-\frac{1}{n}\sum_{c=1}^{\#c} n_c\mu_t^2 \\
&= \frac{1}{n}\sum_{c=1}^{\#c} n_c\mu_{ct}^2-2\mu_t\frac{1}{n}\sum_{c=1}^{\#c} n_c\mu_{ct}-\frac{1}{n}\sum_{c=1}^{\#c} n_c\mu_t^2 \\
&= \frac{1}{n}\sum_{c=1}^{\#c} n_c\left(\mu_{ct}^2-2\mu_t\mu_{ct}-\mu_t^2\right) \\
&= \frac{1}{n}\sum_{c=1}^{\#c} n_c\left(\mu_{ct}-\mu_t\right)^2\equiv\sigma_{t,bet}^2. \qquad (59)
\end{aligned}$$

From Casella and Berger [1990], Theorem 11.2.11, we know that,

$$\begin{aligned}
\sigma^2_{t,bet} = \frac{1}{n}\sum_{c=1}^{\#c} n_c\left(\mu_{ct}-\mu_t\right)^2 &= \frac{1}{n}\sum_{c=1}^{\#c}\sum_{k=1}^{n_c}\left(y_{kct}-\mu_{t,\text{HT}}\right)^2 - \frac{1}{n}\sum_{c=1}^{\#c}\sum_{k=1}^{n_c}\left(y_{kct}-\mu_{ct}\right)^2 \\
&= \sigma^2_t - \sum_{c=1}^{\#c}\frac{n_c}{n}\sum_{k=1}^{n_c}\frac{\left(y_{kct}-\mu_{ct}\right)^2}{n_c} \\
&= \sigma^2_t - \sum_{c=1}^{\#c}\frac{n_c}{n}\sigma^2_{ct} \qquad (60)
\end{aligned}$$

hence, we can interpret $\sigma^2_{t,bet}$ as the between-cluster variance.

Thus, (58) simplifies to

$$\begin{aligned}
\text{Var}(\hat{\mu}_{t,\text{HT}}) &= \mathbb{E}\left(\frac{1}{\#T_t}\right)\left(\sum_{c=1}^{\#c}\frac{n_c}{n}\left(\mu^2_{ct}+\frac{n_c-s_c}{n_c-1}\frac{\sigma^2_{ct}}{s_c}\right)-\mu^2_t\right) \\
&\quad +\mathbb{E}\left(\frac{1}{\#T_t}-1\right)\left(\sum_{c=1}^{\#c}\sum_{c'\neq c}\frac{\pi_{cc'}\mu_{ct}\mu_{c't}}{s(s-1)}+\mu^2_t\right) \\
&= \mathbb{E}\left(\frac{1}{\#T_t}\right)\left(\sigma^2_{t,bet}+\sum_{c=1}^{\#c}\frac{n_c}{n}\left(\frac{n_c-s_c}{n_c-1}\frac{\sigma^2_{ct}}{s_c}\right)\right) \\
&\quad +\mathbb{E}\left(\frac{1}{\#T_t}-1\right)\left(\sum_{c=1}^{\#c}\sum_{c'\neq c}\frac{\pi_{cc'}\mu_{ct}\mu_{c't}}{s(s-1)}+\mu^2_t\right) \qquad (61)
\end{aligned}$$

We derive the covariance term through the property:

$$\text{cov}(\hat{\mu}_{1,\text{HT}},\hat{\mu}_{0,\text{HT}}) = \mathbb{E}(\hat{\mu}_{1,\text{HT}}\hat{\mu}_{0,\text{HT}}) - \mathbb{E}(\hat{\mu}_{1,\text{HT}})\mathbb{E}(\hat{\mu}_{0,\text{HT}}) \qquad (62)$$

Note that:

$$\begin{aligned}
\hat{\mu}_{1,\text{HT}}\hat{\mu}_{0,\text{HT}} &= \left(\sum_{c=1}^{\#c}\frac{S_cT_{c1}}{\#T_1}\sum_{k=1}^{n_c}\frac{y_{kc1}S_{kc}}{s_c}\right)\left(\sum_{c'=1}^{\#c}\frac{S_cT_{c0}}{\#T_0}\sum_{k^*=1}^{n_{c'}}\frac{y_{kc0}S_{kc}}{s_c}\right)\\
&= \left(\sum_{c=1}^{\#c}\sum_{k=1}^{n_c}y_{kc1}\frac{S_cT_{c1}S_{kc}}{\#T_1s_c}\right)\left(\sum_{c'=1}^{\#c}\sum_{k^*=1}^{n_{c'}}y_{k^*c'0}\frac{S_{c'}T_{c'0}S_{k^*c'}}{\#T_0s_{c'}}\right)\\
&= \sum_{c=1}^{\#c}\sum_{k=1}^{n_c}\sum_{c'=1}\sum_{k^*=1}^{n_{c'}}y_{kc1}y_{k^*c'0}\frac{S_cS_{c'}T_{c1}T_{c'0}S_{kc}S_{k^*c'}}{\#T_1\#T_0s_cs_{c'}}\\
&= \sum_{c=1}^{\#c}\sum_{k=1}^{n_c}\sum_{c'\neq c}\sum_{k^*=1}^{n_{c'}}y_{kc1}y_{k^*c'0}\frac{S_cS_{c'}T_{c1}T_{c'0}S_{kc}S_{k^*c'}}{\#T_1\#T_0s_cs_{c'}} \qquad (63)
\end{aligned}$$

The last equality comes from the fact that, since each cluster is given exactly one treatment, $T_{c1}T_{c0} = 0$. From (41), it follows that

$$\begin{aligned}
\mathbb{E}\left(\hat{\mu}_{1,\text{HT}}\hat{\mu}_{0,\text{HT}}\right) &= \mathbb{E}\left(\sum_{c=1}^{\#c}\sum_{k=1}^{n_c}\sum_{c'\neq c}\sum_{k^*=1}^{n_{c'}}y_{kc1}y_{k^*c'0}\frac{S_cS_{c'}T_{c1}T_{c'0}S_{kc}S_{k^*c'}}{\#T_1\#T_0s_cs_{c'}}\right)\\
&= \sum_{c=1}^{\#c}\sum_{k=1}^{n_c}\sum_{c'\neq c}\sum_{k^*=1}^{n_{c'}}y_{kc1}y_{k^*c'0}\mathbb{E}\left(\frac{S_cS_{c'}T_{c1}T_{c'0}S_{kc}S_{k^*c'}}{\#T_1\#T_0s_cs_{c'}}\right)\\
&= \sum_{c=1}^{\#c}\sum_{k=1}^{n_c}\sum_{c'\neq c}\sum_{k^*=1}^{n_{c'}}y_{kc1}y_{k^*c'0}\frac{\pi_{cc'}}{s(s-1)n_cn_{c'}}\\
&= \frac{1}{s(s-1)}\sum_{c=1}^{\#c}\sum_{c'\neq c}\pi_{cc'}\sum_{k=1}^{n_c}\frac{y_{kc1}}{n_c}\sum_{k^*=1}^{n_{c'}}\frac{y_{k^*c'0}}{n_{c'}}\\
&= \frac{1}{s(s-1)}\sum_{c=1}^{\#c}\sum_{c'\neq c}\pi_{cc'}\mu_{c1}\mu_{c'0} \qquad (64)
\end{aligned}$$

Hence:

$$\begin{aligned}
\text{cov}(\hat{\mu}_{1,\text{HT}},\hat{\mu}_{0,\text{HT}}) &= \mathbb{E}(\hat{\mu}_{1,\text{HT}}\hat{\mu}_{0,\text{HT}}) - \mathbb{E}(\hat{\mu}_{1,\text{HT}})\mathbb{E}(\hat{\mu}_{0,\text{HT}})\\
&= \frac{1}{s(s-1)}\sum_{c=1}^{\#c}\sum_{c'\neq c}\pi_{cc'}\mu_{c1}\mu_{c'0} - \mu_1\mu_0 \qquad (65)
\end{aligned}$$

and so:

$$
\begin{aligned}
\text{Var}\left(\hat{\delta}_{\text{HH}}\right) &= \text{Var}\left(\hat{\mu}_{1,\text{HT}}\right) + \text{Var}\left(\hat{\mu}_{0,\text{HT}}\right) - 2\text{cov}(\hat{\mu}_{1,\text{HT}}, \hat{\mu}_{0,\text{HT}}) \\
&= \mathbb{E}\left(\frac{1}{\#T_1}\right)\left(\sigma^2_{1,\text{bet}} + \sum_{c=1}^{\#c}\frac{n_c}{n}\left(\frac{n_c - s_c}{n_c - 1}\frac{\sigma^2_{c1}}{s_c}\right)\right) \\
&\quad + \mathbb{E}\left(\frac{1}{\#T_1} - 1\right)\left(\sum_{c=1}^{\#c}\sum_{c'\neq c}\frac{\pi_{cc'}\mu_{c1}\mu_{c'1}}{s(s-1)} + \mu_1^2\right) \\
&\quad + \mathbb{E}\left(\frac{1}{\#T_0}\right)\left(\sigma^2_{0,\text{bet}} + \sum_{c=1}^{\#c}\frac{n_c}{n}\left(\frac{n_c - s_c}{n_c - 1}\frac{\sigma^2_{c0}}{s_c}\right)\right) \\
&\quad + \mathbb{E}\left(\frac{1}{\#T_0} - 1\right)\left(\sum_{c=1}^{\#c}\sum_{c'\neq c}\frac{\pi_{cc'}\mu_{c0}\mu_{c'0}}{s(s-1)} + \mu_0^2\right) \\
&\quad - \frac{2}{s(s-1)}\sum_{c=1}^{\#c}\sum_{c'\neq c}\pi_{cc'}\mu_{c1}\mu_{c'0} + 2\mu_1\mu_0
\end{aligned} \tag{66}
$$