Adaptive Design of Clustered Experiments

Carlos Gonzalez-Perez, University of Oxford

Harvard's Graduate Workshop in Econometrics

April 13, 2026

Outline

Outline

Motivation

- Surveys and experiments have become **larger**
  - Not just by adding more units, but by spanning **larger geographical regions**
  - Either physically (provinces, schools) or figuratively (markets, industries)
- **Why?** [Muralidharan and Niehaus, 2017]
  - **External Validity.** We want our results to be representative of the population of interest
  - Get better sense of implementation dynamics
  - Governmental support

Motivation

- Ideally, draw units directly from the population of interest, but
  - **Logistically infeasible** and **very costly**
  - Imagine a researcher going all around the country interviewing people
- **Solution?**
  - Run **multistage/stratified** surveys and experiments [Sedgwick, 2015], [Barcaroli et al., 2022]
  - **Structure**
    - Split the population of interest in *independent* clusters
    - Randomly select a number of clusters K
    - Randomly select a set of units N from each of the K clusters
    - Tower the process if needed

Motivation

- **Cluster** related **fixed costs**: partnership with local authorities, travel expenses, additional staff, location specific infrastructure...
- When there is a hard **budget**, a clear **trade-off** emerges
  - **number of clusters** (less correlation between observations hence more efficient estimators but also more expensive)
  - and **number of units** per cluster (higher variance but less expensive)
- This trade-off is mediated by the degree of **within-cluster correlation**
- **Research Question**
  - **Can we design optimal (adaptive) multistage sampling experiments under unknown within-cluster correlation?**

Some Examples

- **Development Economics**
  - [Mbiti et al., 2019] (QJE) randomly sample 10 districts from mainland Tanzania, and then randomly allocate treatment (combination of school grants and teacher incentives) to 35 schools within each district
  - [Karing, 2024] (QJE) select 4 out of Sierra Leone's 14 districts (together with the Ministry of Health and according to some logistic and medical considerations), and then *randomly* allocated treatment (a bracelet system associated to vaccination) to 30 clinics per district
  - [Heß et al., 2021] (REStud) analyze the Gambian CDD program (funds at the village level for community development projects). They selected 88 out of 114 wards and then randomized treatment across villages within wards

Some Examples

- **Survey Design** (prevalent approach)
  - [Sedgwick, 2015] Survey on British public opinion on the use of personal medical data. Randomly select postcodes from the UK (proportionally to size), and then randomly select households from postcodes
  - [Osei and Zhuang, 2020] Survey entrepreneurial women in rural Ghana. Randomly sample 15 communities within a district, and then survey an average of 22 female entrepreneurs per community

Problems in Unit vs Cluster Selection

- **Stratification** seems **prevalent** in RCT and survey designs, but there are **"no clear guidelines"** driving this process
- Selection based on *feel*, logistics, or, even worse, **common misconceptions**
  - Sample **as many clusters as possible**, UN Handbook of Household Sample Surveys in Developing and Transition Economies, [Tam, 2005]
  - **Fix number of clusters** in advance, [Sedgwick, 2015]
  - **Guess** $\hat{\rho}$, do minmax or estimate it from previous studies, [Barcaroli et al., 2022]

Main Contributions

- **Framework** for optimal K vs N selection in stratified surveys and experiments under budget constraints and linear costs
- Characterize the **oracle policy** (and oracle variance) under known correlation
- Propose an **optimal algorithm** with **unknown cluster correlation**
  - Show its excess variance is **negligible** compared to the variance of the oracle
  - Show its excess variance is **unimprovable** through matching lower bounds
  - Show its excess variance is **non-trivial** when compared to optimal static policies
- Algorithm uses a **two-stage approach**
  - **First stage learns correlation** through a stopping algorithm
  - **Second stage** uses first-stage estimates as **plug-ins for optimization**

Outline

Literature Review in Survey Design

- Large literature on cluster vs unit selection in **survey design** [Biemer and Lyberg, 2003], [Bethel, 1989], [Barcaroli et al., 2022], where
  - They emphasize the role of variable and **cluster costs**
  - They emphasize the role of **within-cluster correlation** (ICC)
  - Despite impressive theoretical and algorithmic display featuring complicated objectives, DGP and cost structures, **no attempt at learning ICC**

Literature Review in Economics

- Little econometrics literature on K vs N, but complementary to
  - **covariate** based **stratified treatment allocation** [Cytrynbaum, 2021], [Cytrynbaum, 2024], [Bai, 2022], [Bai et al., 2024]
  - **adaptive experimental design** in economics [Kasy and Sautmann, 2021], [Cesa-Bianchi et al., 2025], [Adusumilli, 2026], [Imbens et al., 2025], [Li et al., 2026]
    - **classic** literature on (panel data) econometrics with **nested random effects** [Wooldridge, 2003]
    - bandit literature on **optimal sampling designs** [Lattimore and Szepesvári, 2020], [Kiefer and Wolfowitz, 1960]
- Use **time-uniform sequences** results in [Howard et al., 2021]

Outline

Model

- Focus on **survey design** problem. Experimental design on its simplest difference-in-means version is completely equivalent
- (Nested) **random effects** model

$$y_{ki} = \mu + \varepsilon_k + \varepsilon_{ki}, \qquad \mathbb{E}[\varepsilon_k] = \mathbb{E}[\varepsilon_{ki}] = 0, \qquad \mathbb{E}[\varepsilon_k^2] = \rho, \;\; \mathbb{E}[\varepsilon_{ki}^2] = 1 - \rho$$

  - $\mathbb{E}[y_{ki}] = \mu$
  - equal unit level variance, i.e. $\mathrm{Var}(y_{ki}) = 1$ for all k, i, (*can be relaxed*)
  - **cross-cluster independence**, i.e. $\mathrm{CoVar}(y_{ki}, y_{lj}) = 0$ for all i, j and $k \neq j$,
  - **bounded within-cluster equicorrelation**, i.e. $\mathrm{CoVar}(y_{ki}, y_{kj}) = \rho$ for all i, j, and all k (*hoping to relax*) with $\boldsymbol{\rho \in [\rho_m, \rho_M]} \subset (0, 1)$

Designer's Problem

- **Estimator**

$$\hat{\mu}(K,N) = \frac{1}{KN}\sum_{k=1}^{K}\sum_{i=1}^{N} y_{ki}$$

- **Minimize $\hat{\mu}$ MSE subject to a budget constraint and linear costs**
  - $\hat{\mu}$ is unbiased under stratified randomization, so simply minimize variance

$$\min_{K\geq 1, N\geq 1} W(K,N;\rho)$$
$$\text{st } FK + VKN \leq B$$

- where
  - $W(K,N;\rho) = \text{Var}(\hat{\mu}(K,N))$
  - F is the **fixed cost** of sampling an additional cluster
  - V is the **variable cost** of sampling an additional unit
  - $B > F + V$ is the hard **budget**

Designer's Problem

- Some simplifications

$$\text{Solve } \mathrm{Var}(\hat{\mu}(K,N)) = \frac{1-\rho}{KN} + \frac{\rho}{K} \qquad \text{Solve } K \text{ from BC, } K^*(N) = \frac{B}{F+VN}$$

- The problem reduces to

$$\min_{1 \le N \le (B-F)/V} \frac{F+VN}{B}\left(\rho + \frac{1-\rho}{N}\right)$$

- Define the **optimal mapping** (via FOC)

$$K^*(r) = \frac{B}{F+VN^*(r)}, \qquad N^*(r) = \min\left\{\max\left\{1, \sqrt{\frac{F}{V}\frac{1-r}{r}}\right\}, \frac{B-F}{V}\right\}$$

Oracle Policy

- Variance of the optimal mapping $\boldsymbol{W^r(r;\rho)} = W(N^*(r), K^*(r); \rho)$
- Variance of the oracle policy is $\boldsymbol{O(1/B)}$

$$\boldsymbol{W^r(\rho;\rho)} = W(N^*(\rho), K^*(\rho); \rho) = \frac{\left(\sqrt{V(1-\rho)} + \sqrt{F\rho}\right)^2}{B}$$

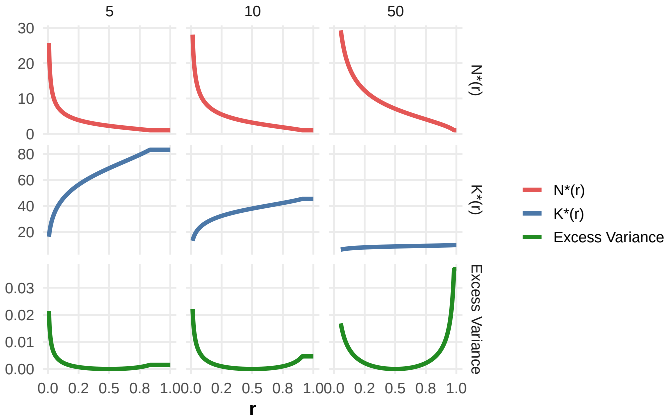
- Define the **local excess variance** of the optimal map as

$$W^r(r,\rho) - W^r(\rho,\rho) \approx \frac{1}{2} W^r_{rr}(\rho;\rho)\ (r-\rho)^2 \sim \boldsymbol{O(1/B)}$$

- More generally, for some policy π define its **excess variance**

$$\boldsymbol{\Delta_\pi} = \mathbb{E}_\pi[W(N_\pi, K_\pi; \rho)] - W^r(\rho;\rho)$$

Comparative Statics of r



$N^*(r)$ (top), $K^*(r)$ (middle) and excess variance $\Delta_\pi(r)$ (bottom) for different F/V ratios.
$B = 1000, V = 2,\ \rho = 0.5$

Outline

$\hat{\rho}$ estimator

- Note
$$\text{Var}(\hat{\mu}_k) = \frac{1 + (N-1)\rho}{N} \implies \rho = \frac{N \ \text{Var}(\hat{\mu}_k) - 1}{N-1}$$
- what suggests the MLE **estimator**

$$\hat{s}^2(K,N) = \frac{1}{K}\sum_{k=1}^{K}(\hat{\mu}_k(K,N) - \hat{\mu}(K,N))^2 \qquad \hat{\rho}(K,N) = \frac{N\hat{s}^2(K,N) - 1}{N-1}$$

- Learning ρ *only* depends on **number of clusters**, not number of units
  - Why? $\text{Var}(\hat{\rho}(K,N))$ is **nearly flat in** N
  - In the paper, I formally characterize the $AVAR$ $\hat{\rho}(K,N)$ under K and N asymptotics

Heuristic Approach

- Say we commit a budget B_0 for a pilot to learn $\hat{\rho}$
  - "Clusters are all that matters" so set N_0 small (≥ 2) and explore as many clusters as budget allows for $K_0 = B_0/(F + VN_0)$
  - Estimate $\hat{\rho}(K_0, N_0)$
- Use remaining budget $B - B_0$ to get as close as possible to $K^*(\hat{\rho}), N^*(\hat{\rho})$, i.e.

$$K_1 = \max\{0, K^*(\hat{\rho}) - K_0\} \qquad \begin{aligned} N_{1k:k\leq K_0} &= \max\{0, N^*(\hat{\rho}) - N_0\} \\ N_{1k:K_0\leq k\leq K_1} &= N^*(\hat{\rho}) \end{aligned}$$

- We won't worry about $N_0 > N^*(\hat{\rho})$, but what if $K_0 > K^*(\hat{\rho})$?
  - We have **oversampled clusters** and incurred in **irreversible fixed costs**. No way of getting close to $K^*(\rho)$
  - Mathematically, the error is **not local** to ρ

Tension in Ideal Two-Stage Protocol

- High B_0 gets us a better $\hat{\rho}(N_0, K_0)$, so $(K^*(\hat{\rho}), N^*(\hat{\rho})) \approx (K^*(\rho), N^*(\rho))$
- But too high B_0 renders the optimal map $(K^*(\hat{\rho}), N^*(\hat{\rho}))$ budget infeasible
- In summary, we want to set K_0 **as large as possible without compromising the second stage optimization**
- **Solution?** Run the initial **pilot through a stopping algorithm** that delays stopping as much as possible, but stops before oversampling $k > K^*(\rho)$ (with high probability)

Outline

Stopping Algorithm

- We want a **stopping alg** $\mathscr{A} : H_{k-1} \mapsto [0, 1]$ which induces a dist over stopping times τ

$$\mathbb{P}_{\mathscr{A}}(\tau \leq K^*(\rho) \;\text{ AND }\; |\, \hat{\rho}(\tau, N_0) - \rho \,| \leq D_\tau(\alpha)) \geq 1 - \alpha \tag{1}$$

Stopping Algorithm

- We want a **stopping alg** $\mathscr{A} : H_{k-1} \mapsto [0,1]$ which induces a dist over stopping times τ

$$\mathbb{P}_{\mathscr{A}}(\tau \leq K^*(\rho) \text{ AND } |\hat{\rho}(\tau, N_0) - \rho| \leq D_\tau(\alpha)) \geq 1 - \alpha \tag{1}$$

- Consider an algorithm which, in every period, constructs a CI $C_k(\alpha) = [\hat{\rho}(k, N_0) - D_k(\alpha), \hat{\rho}(k, N_0) + D_k(\alpha)]$ which **contains** ρ **uniformly with high probability**, i.e.

$$\mathbb{P}\left(|\hat{\rho}(k, N_0) - \rho| \leq D_k(\alpha) \ \forall k\right) \geq 1 - \alpha$$

  - If $k + 1 \leq K^*(\hat{\rho} - D_k(\alpha))$ sample an additional cluster, else stop (provided $K^*(r)$ is increasing in r)
  - This alg satisfies Eq 1 by design (and stops in finite time wp 1 given $B < \infty$)

Stopping Algorithm

- We want a **stopping alg** $\mathcal{A} : H_{k-1} \mapsto [0,1]$ which induces a dist over stopping times τ

$$\mathbb{P}_{\mathcal{A}}(\tau \leq K^*(\rho) \ \text{AND} \ \mid \hat{\rho}(\tau, N_0) - \rho \mid \leq D_\tau(\alpha)) \geq 1 - \alpha \tag{1}$$

- Consider an algorithm which, in every period, constructs a CI $C_k(\alpha) = [\hat{\rho}(k, N_0) - D_k(\alpha), \hat{\rho}(k, N_0) + D_k(\alpha)]$ which **contains** ρ **uniformly with high probability**, i.e.

$$\mathbb{P}\left(\mid \hat{\rho}(k, N_0) - \rho \mid \leq D_k(\alpha) \ \ \forall k\right) \geq 1 - \alpha$$

  - If $k + 1 \leq K^*(\hat{\rho} - D_k(\alpha))$ sample an additional cluster, else stop (provided $K^*(r)$ is increasing in r)
  - This alg satisfies Eq 1 by design (and stops in finite time wp 1 given $B < \infty$)
- Sadly, **frequentist CI do not satisfy** the time uniformity property

Anytime Confidence Sequence [Howard et al., 2021]

- Instead, consider a cluster summary statistic $\tilde{\mu}_k$ and a 3-diff LL $\mathcal{L}(r \mid \tilde{\mu}_k)$ with density $f_r(\tilde{\mu}_k)$ single-peaked at $r = \hat{\rho}$,
  - **Example.** Under normal random effects ε_k and ε_{ki},

$$\hat{\mu}_k \sim N(\mu, \sigma^2(\rho)), \quad \sigma^2(\rho) = N_0 + \frac{1-\rho}{\rho} \qquad \mathcal{L}(r \mid \hat{\mu}_k) = \frac{1}{\sqrt{2\pi\sigma^2(r)}} \exp\left(-\frac{(\hat{\mu}_k - \mu)^2}{2\sigma^2(r)} \right)$$

Anytime Confidence Sequence [Howard et al., 2021]

- Instead, consider a cluster summary statistic $\tilde{\mu}_k$ and a 3-diff LL $\mathcal{L}(r \mid \tilde{\mu}_k)$ with density $f_r(\tilde{\mu}_k)$ single-peaked at $r = \hat{\rho}$,
  - **Example.** Under normal random effects ε_k and ε_{ki},

$$\hat{\mu}_k \sim N(\mu, \sigma^2(\rho)), \quad \sigma^2(\rho) = N_0 + \frac{1-\rho}{\rho} \qquad \mathcal{L}(r \mid \hat{\mu}_k) = \frac{1}{\sqrt{2\pi\sigma^2(r)}} \exp\left(-\frac{(\hat{\mu}_k - \mu)^2}{2\sigma^2(r)}\right)$$

- For smooth prior $\Pi(r)$ with $\Pi(\rho) > 0$, define the **LR/mixing martingale**

$$M_k(r) := \int \prod_{j=1}^{k} \frac{f_\eta(\tilde{\mu}_k)}{f_r(\tilde{\mu}_k)} \Pi(d\eta)$$

Anytime Confidence Sequence [Howard et al., 2021]

- Instead, consider a cluster summary statistic $\tilde{\mu}_k$ and a 3-diff LL $\mathcal{L}(r \mid \tilde{\mu}_k)$ with density $f_r(\tilde{\mu}_k)$ single-peaked at $r = \hat{\rho}$,
  - **Example.** Under normal random effects ε_k and ε_{ki},

$$\hat{\mu}_k \sim N(\mu, \sigma^2(\rho)), \quad \sigma^2(\rho) = N_0 + \frac{1-\rho}{\rho} \qquad \mathcal{L}(r \mid \hat{\mu}_k) = \frac{1}{\sqrt{2\pi\sigma^2(r)}} \exp\left(-\frac{(\hat{\mu}_k - \mu)^2}{2\sigma^2(r)}\right)$$

- For smooth prior $\Pi(r)$ with $\Pi(\rho) > 0$, define the **LR/mixing martingale**

$$M_k(r) := \int \prod_{j=1}^{k} \frac{f_\eta(\tilde{\mu}_k)}{f_r(\tilde{\mu}_k)} \Pi(d\eta)$$

- Consider the **sequence of confidence intervals**

$$C_k(\alpha)\{r : M_k(r) < 1/\alpha\} \qquad \text{where } D_k(\alpha) = \max\{\hat{\rho} - \inf C_k(\alpha), \sup C_k(\alpha) - \hat{\rho}\}$$

Anytime Confidence Sequence

- **Martingale Interpretation**
  - Classical LR intuition. If the data looks more like the mixing (numerator) than it does for a candidate r (denominator), then $M_k(r)$ grows (evidence against r) and it is excluded from $\{r : M_k(r) < 1/\alpha\}$
- **Theorem 5.1. Anytime Confidence Sequences**

$$\mathbb{P}(\rho \in C_k(\alpha) \ \ \forall k) \geq 1 - \alpha$$

proof

Adaptive Cluster Selection Algorithm (ACS)

Algorithm ACS

Input $B, \alpha, N_0, \mathcal{L}(\cdot \mid \cdot), \Pi(\eta), \rho_m$

Initialize $k = 0,\ C_0(\alpha) = [0, 1]$

while $k + 1 \leq \max\{K^*(\rho_m), K^*(\inf C_k(\alpha))\}$ and $k + 1 \leq \frac{B}{F+VN_0}$

explore an additional cluster $k = k + 1$ and **observe** $(y_{ki})_{N_0}$

compute the sufficient statistic $\tilde{\mu}_k$

update $\mathcal{L}(\cdot \mid \tilde{\mu}_k),\ M_k(r),\ C_k(\alpha)$

set $K_0 = k$ and **compute** $\hat{\rho}(K_0, N_0)$

sample $K_1 = \max\{0, K^*(\hat{\rho}) - K_0\}$

$$N_{1k} = N_{1k:k \leq K_0} = \max\{0, N^*(\hat{\rho}) - N_0\},\ \ N_{1k:K_0 \leq k \leq K_1} = N^*(\hat{\rho})$$

Algorithm Analysis in 3 Steps

- Define the **good event** $\mathcal{E} = \{\rho \in C_k(\alpha)\ \forall k\}$, $\mathbb{P}(\mathcal{E}) \geq 1 - \alpha$ (Thm 5.1)

Algorithm Analysis in 3 Steps

- Define the **good event** $\mathcal{E} = \{\rho \in C_k(\alpha)\ \forall k\}$, $\mathbb{P}(\mathcal{E}) \geq 1 - \alpha$ (Thm 5.1)
- **Lemma 5.2. Algorithm Performance**

$$\Delta_{\texttt{ACS}} \mid \mathcal{E} \leq \frac{1}{2} W_{rr}^{r}(\rho, \rho) \mathbb{E}_{\texttt{ACS}}[D_\tau(\alpha)^2 \mid \mathcal{E}\ ] + R_\tau$$

Algorithm Analysis in 3 Steps

- Define the **good event** $\mathcal{E} = \{\rho \in C_k(\alpha)\ \forall k\}$, $\mathbb{P}(\mathcal{E}) \geq 1 - \alpha$ (Thm 5.1)
- **Lemma 5.2. Algorithm Performance**

$$\Delta_{\texttt{ACS}} \mid \mathcal{E} \leq \frac{1}{2} W^r_{rr}(\rho, \rho) \mathbb{E}_{\texttt{ACS}}[D_\tau(\alpha)^2 \mid \mathcal{E}\ ] + R_\tau$$

- **Lemma 5.3. Upper Bound on $D_k(\alpha)$**

$$D_k(\alpha) \leq \sqrt{\frac{2\ln(1/\alpha) + C\ln k}{kI(\hat{\rho}_k)}}(1 + \delta_k) \ \text{ for some seq } \delta_k \to 0$$

Algorithm Analysis in 3 Steps

- Define the **good event** $\mathcal{E} = \{\rho \in C_k(\alpha)\ \forall k\}$, $\mathbb{P}(\mathcal{E}) \geq 1 - \alpha$ (Thm 5.1)
- **Lemma 5.2. Algorithm Performance**

$$\Delta_{\texttt{ACS}} \mid \mathcal{E} \leq \frac{1}{2} W_{rr}^{r}(\rho, \rho) \mathbb{E}_{\texttt{ACS}}[D_\tau(\alpha)^2 \mid \mathcal{E}\ ] + R_\tau$$

- **Lemma 5.3. Upper Bound on $D_k(\alpha)$**

$$D_k(\alpha) \leq \sqrt{\frac{2\ln(1/\alpha) + C \ln k}{k I(\hat{\rho}_k)}}(1 + \delta_k) \ \text{ for some seq } \delta_k \to 0$$

- **Lemma 5.4. Linear Stopping Time**

$$\text{for all sufficiently large } B, \ \ \mathbb{P}_{\texttt{ACS}}(\tau \in \Theta(B) \mid \mathcal{E}\ ) = 1$$

Upper Bound Theorem

Theorem 5.5. Upper Bound on ACS

Under the assumptions in Lemmas 5.2, 5.3 and 5.4 and $0 < I_m < I(r) < \infty$, wp $1 - \alpha$, there exists a constant $C < \infty$ such that

$$\Delta_{\texttt{ACS}} \lesssim C \cdot \frac{\ln(1/\alpha) + \ln B}{B^2}$$

- The excess variance of `ACS` is **asymptotically negligible** compared to the oracle variance $O(1/B) >> O(\ln B/B^2)$

Outline

Intuition for Lower Bound

- Show that `ACS` excess variance is **essentially unimprovable** by deriving **matching lower bounds** on the problem up to logarithmic terms
- **Proof Structure**
  - Consider two DGPs indexed by $\overline{\rho} = \rho + \delta$ and $\underline{\rho} = \rho - \delta$ with $\delta \sim O(B^{-1/2})$
  - δ is **small enough** such that it is difficult to distinguish across instances
  - δ is **large enough** such that misidentification is costly
  - Finally, argue that any policy which does well under $\overline{\rho}$ must not do very well under $\underline{\rho}$ and viceversa

Graphical Representation of Lower Bound

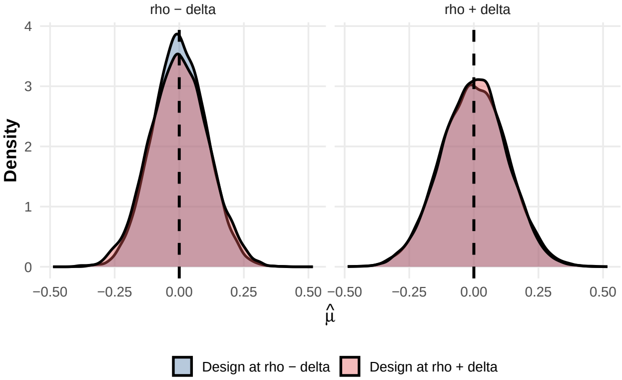


Figure: Distribution of $\hat{\mu}(K, N)$ under $\underline{\rho} = \rho - \delta$ (left) and $\overline{\rho} = \rho + \delta$ (right). $\rho = 0.3$, $\delta = 0.1$, $B = 1000$, $F = 20$, $V = 2$. 10,000 replications.

Lower Bound Theorem

Theorem 6.1. Lower Bound on Excess Variance

For any policy π, such that $P_\pi\left(\sum_k^K (F + VN_k) \leq B\right) = 1$ there exists a DGP indexed by $\rho \in (\rho_m, \rho_M)$ such that

$$\Delta_\pi \gtrsim \frac{1}{B^2}\frac{V}{44\rho}\sqrt{\frac{FV(1-\rho)}{\rho}} \sim O(1/B^2)$$

- `ACS` is **nearly-optimal** up to logarithmic and constant terms

Policy Comparison

- The excess variance of `ACS` $O(\ln B/B^2)$ is **not trivial**
- **Lemma 6.2.**
  - Any static policy which takes a **mispecified** $\tilde{\rho} = \rho + b$,
  - and the **minmax optimal** policy over $\rho \in [\rho_m, \rho_M]$
  - accrue an excess variance of $\boldsymbol{O(1/B)} >> O(\ln B/B^2)$

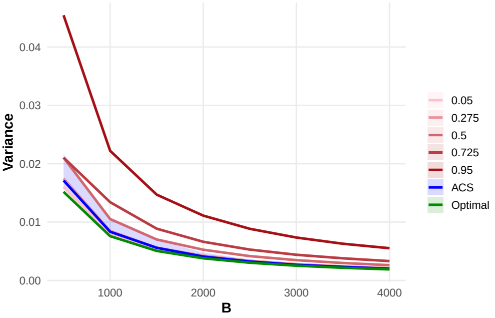
Outline

Simulations

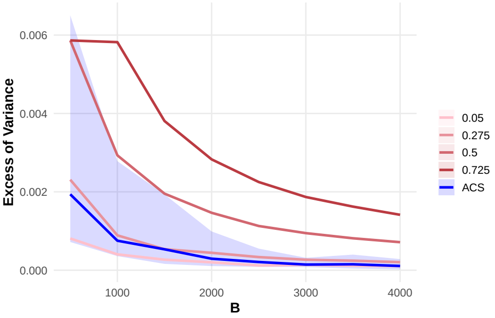
Policy	K	N	Share $K_0 > K^*(\rho)$	Δ_π	$Q_{0.05}(\Delta_\pi)$	$Q_{0.95}(\Delta_\pi)$
Oracle	77	9	0	0	-	-
`ACS`	**77**	**9.5**	**.03**	**.00017**	**.000060**	**.00042**
Fixed .05	60	15	0	.00014	-	-
Fixed .275	100	5	1	.00027	-	-
Fixed .5	115	3	1	.00095	-	-
Fixed .725	125	2	1	.0019	-	-
Fixed .95	136	1	1	.0048	-	-

Table: Performance Comparison across Policies $\rho = 0.1$. $\alpha = 0.05$, $B = 3000$, $F = 20$, $V = 2$. 100 replications.

Simulations: Performance as a Function of B



(a) Variance $W(N_{\texttt{ACS},K_{\texttt{ACS};\rho}})$



(b) excess variance $W(N_{\texttt{ACS},K_{\texttt{ACS};\rho})-W^{r}(\rho,\rho)}$

Figure: Variance (left) and Excess Variance (right) as a Function of B across Policies. $\rho = 0.1$, $\alpha = 0.05$, $F = 20$, $V = 2$. 100 replications.

Simulations: Selection of K_0

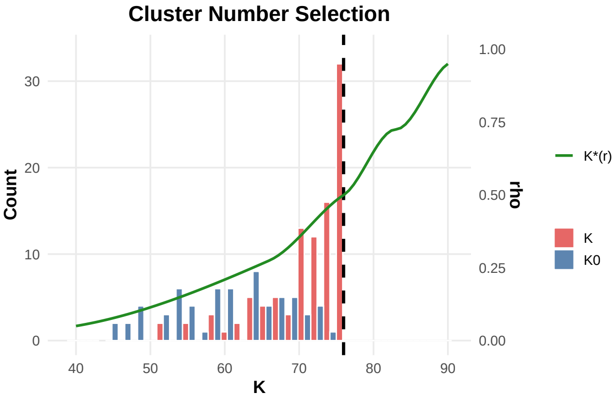


Figure: $K_{0\texttt{ACS}}$ and $K_{\texttt{ACS}}$ (histogram left), and $K^*(\rho)$ (line plot right). $\rho = 0.5$, $\alpha = 0.05$, $B = 2000$, $F = 20$, $V = 2$. Dashed line equals $K^*(0.5)$.

Outline

Conclusion

- Proposed a **framework** for optimal K vs N selection in stratified surveys and experiments under budget constraints and linear costs
- Characterized the **oracle policy** (and oracle variance) under known correlation
- Proposed an **optimal algorithm** with **unknown cluster correlation**
  - Showed its excess variance is **negligible** compared to the variance of the oracle
  - Showed its excess variance is **unimprovable** through matching lower bounds
  - Showed its excess variance is **non-trivial** when compared to optimal static policies

Thank you!

carlos.gonzalezperez@economics.ox.ac.uk
carlosgonzalezperez@fas.harvard.edu

Bibliography I

Adusumilli, K. (2026).
How to sample and when to stop sampling: The generalised wald problem and minimax policies.
Review of Economic Studies, 93(1):1–34.

Bai, Y. (2022).
Optimality of matched-pair designs in randomized controlled trials.
American Economic Review, 112(12):3911–3940.

Bai, Y., Jiang, L., Romano, J. P., Shaikh, A. M., and Zhang, Y. (2024).
Covariate adjustment in experiments with matched pairs.
Journal of Econometrics, 241(1):105740.

Barcaroli, G., Fasulo, A., Guandalini, A., and Terribili, M. D. (2022).
Two-stage sampling design and sample selection with the r package r2beat.
arXiv preprint arXiv:2209.06532.

Bibliography II

Bethel, J. (1989).
Sample allocation in multivariate surveys.
Survey methodology, 15(1):47–57.

Biemer, P. P. and Lyberg, L. E. (2003).
Introduction to survey quality.
John Wiley & Sons.

Cesa-Bianchi, N., Colomboni, R., and Kasy, M. (2025).
Adaptive maximization of social welfare.
Econometrica, 93(3):1073–1104.

Cytrynbaum, M. (2021).
Optimal stratification of survey experiments.
arXiv preprint arXiv:2111.08157.

Bibliography III

Cytrynbaum, M. (2024).
Covariate adjustment in stratified experiments.
Quantitative Economics, 15(4):971–998.

Heß, S., Jaimovich, D., and Schündeln, M. (2021).
Development projects and economic networks: Lessons from rural gambia.
The Review of Economic Studies, 88(3):1347–1384.

Howard, S. R., Ramdas, A., McAuliffe, J., and Sekhon, J. (2021).
Time-uniform, nonparametric, nonasymptotic confidence sequences.
The Annals of Statistics, 49(2):1055–1080.

Imbens, G., Qin, C., and Wager, S. (2025).
Admissibility of completely randomized trials: A large-deviation approach.
arXiv preprint arXiv:2506.05329.

Bibliography IV

Karing, A. (2024).
Social signaling and childhood immunization: A field experiment in sierra leone.
The Quarterly Journal of Economics, 139(4):2083–2133.

Kasy, M. and Sautmann, A. (2021).
Adaptive treatment assignment in experiments for policy choice.
Econometrica, 89(1):113–132.

Kiefer, J. and Wolfowitz, J. (1960).
The equivalence of two extremum problems.
Canadian Journal of Mathematics, 12:363–366.

Lattimore, T. and Szepesvári, C. (2020).
Bandit algorithms.
Cambridge University Press.

Bibliography V

- Li, D., Raymond, L., and Bergman, P. (2026).
  Hiring as exploration.
  Review of Economic Studies, 93(2):1200–1240.

- Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., and Rajani, R. (2019).
  Inputs, incentives, and complementarities in education: Experimental evidence from tanzania.
  The Quarterly journal of economics, 134(3):1627–1673.

- Muralidharan, K. and Niehaus, P. (2017).
  Experimentation at scale.
  Journal of Economic Perspectives, 31(4):103–124.

Bibliography VI

Osei, C. D. and Zhuang, J. (2020).
Rural poverty alleviation strategies and social capital link: The mediation role of women entrepreneurship and social innovation.
Sage Open, 10(2):2158244020925504.

Sedgwick, P. (2015).
Multistage sampling.
Bmj, 351.

Tam, B. (2005).
Household sample surveys in developing and transition countries.

Wooldridge, J. M. (2003).
Cluster-sample methods in applied econometrics.
American Economic Review, 93(2):133–138.

Proof of Anytime Confidence Sequence

- By (simplified) Ville's Inequality, for any non-negative martingale M_k and any real $\alpha > 0$

$$\mathbb{P}\left(\exists\, k : M_k \geq \frac{1}{\alpha}\right) \leq \alpha \mathbb{E}[M_0]$$

- Thus, it remains to show that M_k is non-negative martingale with $\mathbb{E}[M_0] = 1$

$$\begin{aligned}
\mathbb{E}_r\left[\prod_k \frac{f_\eta(\tilde{\mu}_k)}{f_r(\tilde{\mu}_k)}\right] &= \prod_k \mathbb{E}_r\left[\frac{f_\eta(\tilde{\mu}_k)}{f_r(\tilde{\mu}_k)}\right] && \text{cross-cluster independence} \\
&= \prod_k \int \frac{f_\eta(\tilde{\mu}_k)}{f_r(\tilde{\mu}_k)} f_r(\tilde{\mu}_k) d\tilde{\mu}_k && \text{definition of } \mathbb{E}_r \\
&= \prod_k \int f_\eta(\tilde{\mu}_k) d\tilde{\mu}_k = 1 && \text{valid density}
\end{aligned}$$

Proof of Anytime Confidence Sequence

- Therefore,

$$\mathbb{E}_r\left[\int \prod_k \frac{f_\eta(\tilde{\mu}_k)}{f_r(\tilde{\mu}_k)} \Pi(d\eta)\right] \overset{\text{Fub}}{=} \int \mathbb{E}_r\left[\prod_k \frac{f_\eta(\tilde{\mu}_k)}{f_r(\tilde{\mu}_k)}\right] \Pi(d\eta) = \int 1 \Pi(d\eta) = 1$$

return