Adaptive Design of Clustered Experiments

Carlos Gonzalez-Perez, University of Oxford

UK-GEN Warwick

June 13th, 2026

Outline

Outline

Hello!

- Carlos Gonzalez-Perez, **JMC** in Econometrics at Oxford
- **Research**
  - Adaptive Experimental Design
  - Online Learning and Bandits
  - Micro Theory on Learning Problems
- website: presidente-carlos.github.io

Outline

Motivation

- Surveys and experiments have become **larger**
  - both in units and **scope**
  - spanning across several **clusters** like districts, provinces, schools, markets
  - **Why?** External validity and efficiency [Muralidharan and Niehaus, 2017]
- **Key feature** of clustered designs: **number of clusters** K vs **number of units per cluster** N (with randomization at every stage)
  - $(N_k)_k$ with $K = \sum_k \mathbb{1}(N_k > 0)$

Motivation

- **Statistically:** If designer believes **observations within clusters are more correlated**, then, the more clusters, the better!
- **Costs:** In practice, sampling from different clusters is **expensive**
  - partnership with local authorities, travel expenses, additional staff, location specific infrastructure...
- Cost-**optimal design** balances statistical accuracy and costs
  - **Key parameter—ICC:** “how correlated are units within clusters”
- **Research Questions?**
  - **Can we design optimal clustered experiments under unknown ICC?**
  - **Can we do so without compromising inference?**

Contributions

- **Framework** for optimal design of clustered experiments with BC
- **Two adaptive algorithms** that implement a **near-optimal** design without knowing the ICC
  - induce estimators with **negligible** excess variance compared to the oracle
  - essentially **unimprovable** and **outperform any misspecified static policy**
- Under a symmetry assumption, **adaptive data collection based on even moments of the data does not affect inference**
  - my algorithms sample based on sample variances (which are asymptotically even), so the induced estimators remain asymptotically normal

Outline

Model

- **Today—Survey Problem**
- Nested **random effects** model

$$y_{ki} = \mu + u_k + \varepsilon_{ki}, \qquad \mathbb{E}[u_k] = \mathbb{E}[\varepsilon_{ki}] = 0, \qquad \mathbb{E}[u_k^2] = \rho \;\; \mathbb{E}[\varepsilon_{ki}^2] = \sigma, \qquad u_k \perp\!\!\!\perp \varepsilon_{ki}$$

  - Define $\lambda = \sigma/\rho$ as the **variance ratio**, ICC $= \frac{1}{1+\lambda}$
  - **[A1]—Compactness:** Model is parameterized by $\omega = (\rho, \lambda) \in \Omega = [\rho_m, \rho_M] \times [\lambda_m \times \lambda_M]$ with Ω **known** and $\rho_m > 0$

Designer's Problem

- **Estimator**

$$\hat{\mu}(K,N) = \frac{1}{KN}\sum_{k=1}^{K}\sum_{i=1}^{N} y_{ki}, \qquad \mathbb{E}[\hat{\mu}] = \mu, \qquad \mathrm{Var}(\hat{\mu}) = W(K,N,\omega) = \frac{\rho}{K}\left(1 + \frac{\lambda}{N}\right)$$

- **Goal: Minimize $\hat{\mu}$ MSE** (variance) subject to **BC with linear costs**

$$\min_{K\geq 2, N\geq 2} W(K,N,\omega) \qquad \text{st } FK + VKN \leq B$$

  - K, N boundaries for estimation purposes
  - continuous relaxation of the problem

Optimal Map and Oracle Variance

- **Optimal map** (OM) is only a function of λ and sensitive to misspecification

$$K^*(\lambda) = \frac{B}{F + VN^*(\omega)}, \qquad N^*(\lambda) = \min\left\{\max\left\{2, \sqrt{\lambda\frac{F}{V}}\right\}, \frac{B-2F}{2V}\right\}$$

Optimal Map and Oracle Variance

- **Optimal map** (OM) is only a function of λ and sensitive to misspecification

$$K^*(\lambda) = \frac{B}{F + VN^*(\omega)}, \qquad N^*(\lambda) = \min\left\{\max\left\{2, \sqrt{\lambda\frac{F}{V}}\right\}, \frac{B-2F}{2V}\right\}$$

- **[A2]—Interior Solution:** $N^*(\lambda) \in (2, (B-2F)/(2V))\ \forall\lambda \in [\lambda_m, \lambda_M]$

Optimal Map and Oracle Variance

- **Optimal map** (OM) is only a function of λ and sensitive to misspecification

$$K^*(\lambda) = \frac{B}{F + VN^*(\omega)}, \qquad N^*(\lambda) = \min\left\{\max\left\{2, \sqrt{\lambda\frac{F}{V}}\right\}, \frac{B-2F}{2V}\right\}$$

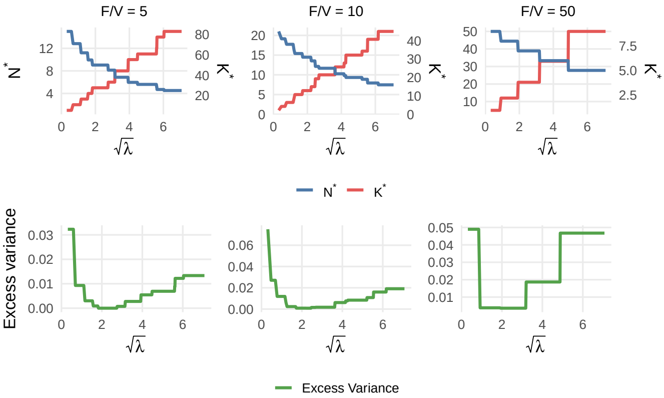
- **[A2]—Interior Solution:** $N^*(\lambda) \in (2, (B-2F)/(2V))\ \forall\lambda \in [\lambda_m, \lambda_M]$
- For $\omega = \omega_0$

Oracle Variance: $W(K^*(\lambda_0), N^*(\lambda_0), \omega_0) = \dfrac{\rho_0\left(\sqrt{V\lambda_0} + \sqrt{F}\right)^2}{B} \sim \boldsymbol{O(1/B)}$

Excess Variance: $\Delta_\pi = W(K^\pi, N^\pi, \omega_0) - W(K^*(\lambda_0), N^*(\lambda_0), \omega_0)$

Excess Variance OM : $W(K^*(\lambda), N^*(\lambda), \omega_0) - W(K^*(\lambda_0), N^*(\lambda_0), \omega_0) \approx \dfrac{C}{B}(\lambda - \lambda_0)^2$

λ Misspecification



$N^*(\lambda)$ (top-left), $K^*(\lambda)$ (top-right) and EVOM (bottom) for different F/V ratios. $\mu = 0$, $\omega_0 = (1, 5)$, $B = 1000, V = 2$

Outline

An Optimal Algorithm in a Nutshell

- **Explore-then-Commit** logic
  - Run a small pilot (K_p^π, N_p^π) to learn λ, i.e. $\hat{\lambda}_p = \hat{\lambda}(K_p^\pi, N_p^\pi)$
  - Implement the OM design: $(K^\pi, N^\pi) = (K^*(\hat{\lambda}_p), N^*(\hat{\lambda}_p))$
- How to select (K_p^π, N_p^π)?
  - **Problem:** If $K_p^\pi \geq K^*(\hat{\lambda}_p)$ or $N_p^\pi \geq N^*(\hat{\lambda}_p)$, OM is not implementable
  - **Solution:** Set $K_p^\pi = K^*(\lambda_M)$ and $N_p^\pi = N^*(\lambda_m)$
    - By [A1-A2], every OM is implementable

Adaptive Pilot

- `EtC-CS` achieves **negligible excess variance**
  - Easy to implement (only two stages), easy to understand
- Can we do better?
  - Imagine $K^*(\lambda) \in [20, 60]$
  - $K_p^{\texttt{EtC-CS}} = 20,\ K^*(\hat{\lambda}(20, N_p)) = 50 >> 20$
  - Why commit to 50 straightway? Improve $\hat{\lambda}$ by exploring a few more clusters without risk of oversampling
- Sequential cluster sampling in a pilot

$$\tau := \inf\{k : k + 1 > K^*(\hat{\lambda}(k, N_p))\} \wedge \left\lfloor \frac{B}{F + VN_p} \right\rfloor$$

Upper Bounds

- **[A3]—Subgaussianity:** u_k is $\sqrt{\kappa_u}$-subgaussian and ε_{ki} is $\sqrt{\kappa_\varepsilon}$-subgaussian

Theorem 5.5. Upper Bound on EtC-CS and ACS

Under [A1-A3], there exists constants $C_1, C_2 < \infty, \perp\!\!\!\perp B$ such that

$$\mathbb{E}_{\texttt{EtC-CS}}[\Delta_{\texttt{EtC-CS}}] \leq C_2 \frac{\ln(B)}{B^2} \qquad \mathbb{E}_{\texttt{ACS}}[\Delta_{\texttt{ACS}}] \leq C_2 \frac{\ln(B)}{B^2}$$

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

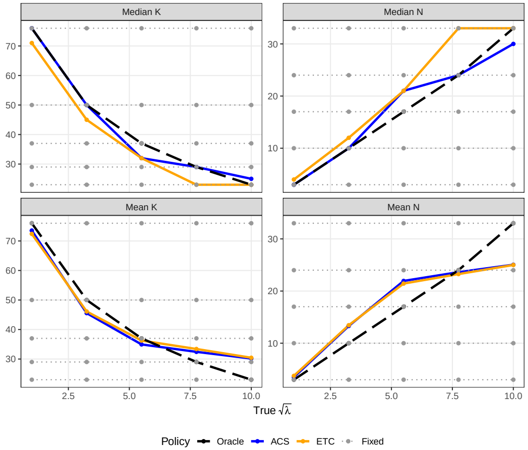
Proof in a Slide

- First: **high-probability bounds**
  - Both `EtC-CS` and `ACS` implement OM so $\Delta_\pi \approx \frac{1}{B}(\hat{\lambda}_p - \lambda_0)^2$
  - Bound $\mid \hat{\lambda}_p - \lambda_0 \mid$ in high probability
    - Decompose $\hat{\lambda}$ as an asymptotic function of martingales with size K_p^π
    - Apply **maximal Bernstein** ineq for each martingale over a **window** $O_B(1)$ provided $K_p^\pi \in [K^*(\lambda_M), K^*(\lambda_m)]$
    - Recover $\mathbb{P}_\pi \left(\Delta_\pi \leq C \, \frac{\ln(1/\alpha)}{B^2}\right) \geq 1 - \alpha$
- $\mathbb{E}_\pi[\Delta_\pi] \leq (1-\alpha) \cdot C \, \frac{\ln(1/\alpha)}{B^2} + \alpha \cdot c$ for all α
- Optimizing wrt to α leads to $\mathbb{E}_\pi[\Delta_\pi] \leq O\left(\frac{\ln B}{B^2}\right)$

Final Comments on Algorithm Performance

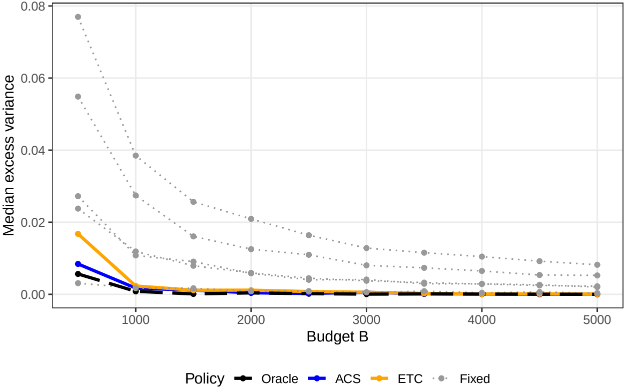
- Despite sharing the same rate, `ACS` **outperforms** `EtC-CS` in simulations
- The paper further shows that
  - These algorithms are **essentially unimprovable** by showing **matching lower bounds** on the problem of $O(1/B^2)$
  - Any **misspecified static policy** (including the minmax optimal) will accrue **excess variance** of $O(1/B) >> O(\ln B/B^2)$, so the rate is **non-trivial**

Simulations—Design



Median (top) and Mean (bottom) (K, N) across policies and λ_0. Fixed policies set at $\lambda \approx (1, 10, 30, 60, 100)$, $\mu = 0$, $\rho_0 = 1$. $B = 2000,\ F = 20,\ V = 2$

Simulations—Excess Variance across B



Median Excess Variance across policies and Budget B. Fixed policies set at $\lambda \approx (1, 10, 30, 60, 100)$, $\mu = 0, \omega_0 = (1, 5)$. $B = 2000$, $F = 20$, $V = 2$

Outline

Inference Problem

- (K^π, N^π) is a function of the **pilot data**, therefore is not obvious that $\hat{\mu}(K^\pi, N^\pi) \mid (K^\pi, N^\pi)$ is unbiased nor asymptotically normal
- Folk knowledge: **When sampling** (the design) **depends on averages, inference is seriously compromised** [Nie et al., 2018], [Zhang et al., 2020], [Hadad et al., 2021]
- What if sampling depends on **even moments of data**?
- I show that under **error symmetry**, group **sample averages are asymptotically independent** of data collection protocols depending on even moments
  - `EtC-CS` and `ACS` designs do not affect inference

Symmetry

- **Why Symmetry?**
  - $\text{Cov}(\hat{\mu}, \hat{s}^2) = \frac{\mathbb{E}[\varepsilon_i^3]}{N} \implies$ symmetry kills the covariance
  - If we can get a CLT to go through $(\hat{\mu}, \hat{s}^2)$, they will be jointly normal
  - Zero-covariance implies independence under joint normality
  - In practice more difficult as policies depend on **random sequences** of $\hat{s}^2$
- **[A4]—Symmetry:** Let $(u_k, \varepsilon_{k1}, \varepsilon_{k2}, \ldots) \stackrel{d}{=} -(u_k, \varepsilon_{k1}, \varepsilon_{k2}, \ldots)$
- **[A5]—Lindeberg Condition**

Applied Theorem

Theorem 6.2. Design Robust Inference

Let $\mathcal{G}$ be the σ-algebra generated by even moments of the full data. Let (K^π, N^π) be $\mathcal{G}$-measurable. Define $\hat{\theta} = \hat{\rho}(1 + \hat{\lambda}/N^\pi)$. Then, under [A1, A4-A5]

$$\frac{\sqrt{K^\pi}\left(\hat{\mu}(K^\pi, N^\pi) - \mu\right)}{\sqrt{\hat{\theta}}} \underset{B}{\to} Z \sim N(0,1) \quad \mathcal{G}\text{-stably with } Z \perp\!\!\!\perp \mathcal{G}$$

- The paper **generalizes** to (i) inference on **group sample averages** $(\hat{\mu}_k)_k$, (ii) any sequential sampling policy, (iii) only asymptotically measurable wrt even σ-algebra
- **Conjecture:** Without symmetry, $\hat{\mu}$ remains asymp normal with bias $\propto \mathbb{E}[u_k^3]/\sqrt{K}$

Proof in a Slide for Applied Theorem

- $\hat{\lambda}$ is asymp an even function of centered errors L_k. Define $\mathcal{G} = \sigma((L_k)_{k\geq 1})$
- Let $\hat{\mu}(K,N) - \mu = \frac{1}{K}\sum_k M_k$, under symmetry $\mathbb{E}[M_k \mid L_k] = 0$
- **Stable CLT** (Corollary 3.1. in [Hall and Heyde, 1980])
  - **Martingale Difference:** Define $\mathcal{F}_k = \sigma((L_j)_{j\geq 1}, M_1, \ldots, M_k)$ (we **frontload** the even errors!), so $(M_k, \mathcal{F}_k)$ is a martingale difference sequence
  - Convergence in prob of the **Conditional Variance** $\frac{1}{K}\sum_k \mathbb{E}[M_k^2 \mid \mathcal{F}_{k-1}] \xrightarrow{p} \theta$
  - **Lindeberg Condition** (vanishing differences)
- SC with **non-random conditional variance** implies **mixing-convergence** and MC implies **independence** of the converging limits

Simulations—Asymptotic Normality

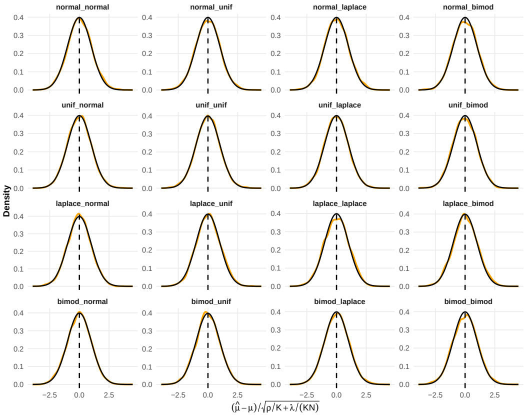


Figure: Empirical pdf for Modal Design Conditional Sample Average $(\hat{\mu}(K^{\pi}, N^{\pi}) - \mu)/\sqrt{\rho/K^{\pi}\,(1+\lambda/N^{\pi})}$ across symmetric distributions. u_k (left), ε_{ik} (right). 10,000 simulations. $\mu = 0, \omega_0 = (1, 5)$. $B = 2000$, $F = 20$, $V = 2$

Outline

Conclusion

- **Framework** for optimal design of clustered experiments with BC
- **Two adaptive algorithms** that implement a **near-optimal** design without knowing the ICC
  - induce estimators with **negligible** excess variance compared to the oracle
  - essentially **unimprovable** and **outperform any misspecified static policy**
- Under a symmetry assumption, **adaptive data collection based on even moments of the data do not affect inference**
  - my algorithms sample based on sample variances (which are asymptotically even), so the induced estimators remain asymptotically normal

Thank you!

carlos.gonzalezperez@economics.ox.ac.uk

carlosgonzalezperez@fas.harvard.edu

Bibliography I

Hadad, V., Hirshberg, D. A., Zhan, R., Wager, S., and Athey, S. (2021).
Confidence intervals for policy evaluation in adaptive experiments.
Proceedings of the national academy of sciences, 118(15):e2014602118.

Hall, P. and Heyde, C. C. (1980).
Martingale limit theory and its application.
Academic press.

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

Nie, X., Tian, X., Taylor, J., and Zou, J. (2018).
Why adaptively collected data have negative bias and how to correct for it.
In *International Conference on Artificial Intelligence and Statistics*, pages 1261–1269. PMLR.

Bibliography II

Zhang, K., Janson, L., and Murphy, S. (2020).
Inference for batched bandits.
Advances in neural information processing systems, 33:9818–9829.

Explore-then-Commit Cluster Selection (EtC-CS)

Algorithm EtC-CS

input $B, F, V, \lambda_m, \lambda_M$

sample $K_p^{\texttt{EtC-CS}} = K^*(\lambda_M)$ clusters and $N_p^{\texttt{EtC-CS}} = N^*(\lambda_m)$ units per cluster

recover $\hat{\lambda}_p = \hat{\lambda}(K_p, N_p)$

implement $K^{\texttt{EtC-CS}} = K^*(\hat{\lambda}_p)$ and $N^{\texttt{EtC-CS}} = N^*(\hat{\lambda}_p)$

Adaptive Cluster Selection Algorithm (ACS)

Algorithm ACS

input $B, F, V, \lambda_m, \lambda_M$

initialize $k = 0$, $N_p^{\texttt{ACS}} = N^*(\lambda_m)$, $\hat{\lambda}_0 = \lambda_m$

while $k + 1 \leq \min\left\{K^*(\hat{\lambda}(k, N_p^{\texttt{ACS}})), \left\lfloor \frac{B}{F+V N_p^{\texttt{ACS}}} \right\rfloor\right\}$

  explore an additional cluster k and **observe** $(y_{ki})_{i=1}^{N_p^{\texttt{ACS}}}$

  update $k = k + 1$

set $K = k$

recover $\hat{\lambda}_{\texttt{ACS}} = \hat{\lambda}(K - 1, N_p^{\texttt{ACS}})$ using $((y_{ki})_{i=1}^{N_p^{\texttt{ACS}}})_{k=1}^{K-1}$

implement $K^{\texttt{ACS}} = K^*(\hat{\lambda}_{\texttt{ACS}})$ and $N^{\texttt{ACS}} = N^*(\hat{\lambda}_{\texttt{ACS}})$