Empirical Methods for Policy Evaluation

Matteo Bobba

Toulouse School of Economics (TSE)

TSE PhD Program (MRes)
Fall 2025

Randomized Experiments (5 Classes)

1. Statistical analysis of randomized experiments (RCTs)
   - SUTVA and assignment mechanisms
   - Randomization inference
   - Randomization designs
   - Stratified RCTs and Clustered RCTs

2. Applications

   ⇒ Risk sharing and seasonal migration (Meghir et al, ReStud 2022)

   ⇒ The production of human capital in the early years (Attanasio et al, AER 2020)

SUTVA and assignment mechanisms

Causal Inference as a Missing Data Problem

- Population of units, indexed by $i = 1, ..., N$
- Treatment indicator W_i taking values 0 and 1
- For $i \in \{1, ..., N\}$ $\exists$ one realized outcome and one missing potential outcome

$$Y_i^{\text{obs}} = Y_i(W_i) = \begin{cases} Y_i(0) \text{ if } W_i = 0 \\ Y_i(1) \text{ if } W_i = 1 \end{cases}$$

$$Y_i^{\text{miss}} = Y_i(1 - W_i) = \begin{cases} Y_i(1) \text{ if } W_i = 0 \\ Y_i(0) \text{ if } W_i = 1 \end{cases}$$

$\Rightarrow$ Unit-level causal effect $Y_i(1) - Y_i(0)$ is unobserved

The Stable Unit Treatment Value Assumption (SUTVA)

- We will generally need to predict, or impute, the missing potential outcome

$\Rightarrow$ Denote $W_{-i} = (W_j)_{j \neq i}$, SUTVA requires that

$$(Y_i(1), Y_i(0)) \perp W_{-i}$$

1. No interference

   $\Rightarrow$ Potential outcomes for unit i do not vary with the treat assigned to any other j

2. Scale Invariance

   $\Rightarrow$ There are no different forms of each treat, which lead to $\neq$ potential outcomes

Sources of Possible Violations of SUTVA

1. Spillovers and equilibrium effects
   - ⇒ Fertilizer in one plot may affect yields in contiguous plots
   - ⇒ Immunization efficacy may depend on the number of people immunized
   - ⇒ Prob(job) after training may be affected by the number of people trained
2. Hidden variations of treatments
   - ⇒ Endogenous compliance to treatment assignment
   - ⇒ Unobserved differences in the method of administering the treatment

Assignment Mechanism

- N units, set $\mathbb{W} = \{0,1\}^N$ of N-vectors with all elements equal to 0 or 1

$\Rightarrow$ The assignment mechanism is a function $P(\mathbf{W}|\mathbf{Y}(0), \mathbf{Y}(1)) \in [0,1]$ such that

$$\sum_{\mathbf{W}\in\{0,1\}^N} P(\mathbf{W}|\mathbf{Y}(0), \mathbf{Y}(1)) = 1$$

$\Rightarrow$ The unit-level assignment probability is

$$p_i(\mathbf{Y}(0), \mathbf{Y}(1)) = \sum_{\mathbf{W}:W_i=1} P(\mathbf{W}|\mathbf{Y}(0), \mathbf{Y}(1))$$

$\Rightarrow$ The propensity score is

$$e(x) = \frac{1}{N(x)} \sum_{i:X_i=x} p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$$

Example of Assignment Mechanism 1

- Two units, and so (2^2) possible values for $\mathbf{W}$

$$\mathbf{W} \in \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

$\Rightarrow$ The assignment mechanism is equal to $P(\mathbf{W}|\mathbf{Y}(0), \mathbf{Y}(1)) = \frac{1}{4}$

$\Rightarrow$ Unit assignment probability $p_i = p = \frac{1}{2}$ for $i = 1, 2$

$\Rightarrow$ No covariates, and so propensity score $e = p = \frac{1}{2}$

Example of Assignment Mechanism 2

- Two units where only assignments with one treated and one control unit

$\Rightarrow$ The assignment mechanism is

$$P(\mathbf{W}|\mathbf{Y}(0),\mathbf{Y}(1)) = \begin{cases} 1/2 & \text{if } \mathbf{W} \in \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\} \\ 0 & \text{if } \mathbf{W} \in \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} \end{cases}$$

$\Rightarrow$ As before, $e = p = \frac{1}{2}$

Regular Assignment Mechanism

1. Individualistic

$$p_i(\mathbf{Y}(0), \mathbf{Y}(1)) = q(Y_i(0), Y_i(1)),\ q(\cdot) \in [0, 1]$$

2. Probabilistic

$$0 < p_i(\mathbf{Y}(0), \mathbf{Y}(1)) < 1$$

3. Unconfounded

$$P(\mathbf{W}|\mathbf{Y}(0), \mathbf{Y}(1)) = P(\mathbf{W}|\mathbf{Y}'(0), \mathbf{Y}'(1)) = P(\mathbf{W})$$

Randomized Experiment

- Individualistic + unconfounded implies

$$P(\mathbf{Y}(0), \mathbf{Y}(1)) = c \cdot \prod_{i=1}^{N} q^{W_i}(1-q)^{1-W_i}$$

- The constant c ensures that the probabilities add to unity
- In this case, $e(x) = p_i(x) = q(x)$

$\Rightarrow$ A regular assignment mechanism with known $q(\cdot)$ is a randomized experiment

Randomization Inference

Two Approaches for Inference (Not Just in RCTs)

1. Asymptotic inference
   - ⇒ (semi-) Parametric models for the conditional mean of observed outcomes
   - ⇒ Observed outcomes vary through random sampling from a population of units
   - ⇒ Compute the distribution of the test statistic through large-sample properties

2. Randomization inference (aka Fisher's exact p-values)
   - ⇒ Nonparametric (no restrictions on the distribution of the potential outcomes)
   - ⇒ Treatment assignments are the sole source of randomness
   - ⇒ The assignment mechanism determines the distribution of the test statistic

Illustrative Example

Unit	Potential Outcomes				
	Cough Frequency (cfa)		Observed Variables		
	$Y_i(0)$	$Y_i(1)$	W_i	X_i (cfp)	Y_i^{obs} (cfa)
1	?	3	1	4	3
2	?	5	1	6	5
3	?	0	1	4	0
4	4	?	0	4	4
5	0	?	0	1	0
6	1	?	0	5	1

The Sharp Null

$$H_0 \ : \ Y_i(0) = Y_i(1) \ \forall i = 1, ..., 6.$$

$\Rightarrow$ The average null hypothesis is weaker than the sharp null hypothesis

- Under H_0, all missings in potential outcomes inferred from obs. outcomes
- One test statistics from example above is

$$\begin{aligned} T(W, Y^{\text{obs}}) &= \mid \overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} \mid \\ &= \mid (Y_1^{\text{obs}} + Y_2^{\text{obs}} + Y_3^{\text{obs}})/3 - (Y_4^{\text{obs}} + Y_5^{\text{obs}} + Y_6^{\text{obs}})/3 \mid \\ &= \mid 8/3 - 5/3 \mid = 1.00 \end{aligned}$$

Treatment Assignments Generate Test Distribution

- We can re-do this for $\binom{6}{3} = 20$ permutations of treatment assignments

$\Rightarrow$ E.g. instead of $W^{\text{obs}} = (1,1,1,0,0,0)$ take $\tilde{W} = (0,1,1,0,0,1)$

- No change in observed outcomes since $Y_i(0) = Y_i(1) = Y_i^{\text{obs}}$ under H_0

- The value of the test statistic for $\tilde{W}$ is

$$\begin{aligned} T(\tilde{W}, Y^{\text{obs}}) &= \mid (Y_2^{\text{obs}} + Y_3^{\text{obs}} + Y_6^{\text{obs}})/3 - (Y_1^{\text{obs}} + Y_4^{\text{obs}} + Y_5^{\text{obs}})/3 \mid \\ &= \mid 6/3 - 7/3 \mid = 0.33 \end{aligned}$$

Computation of Exact p-values

- Calculate the value of the statistic for each assignment vector
- In previous example, each assignment vector has prior probability=1/20
- How unusual is $T(W, Y^{\text{obs}}) = 1.00$ under the sharp null hypothesis?

$\Rightarrow$ There are 16/20 assignments with $T(\tilde{W}, Y^{\text{obs}}) > T(W, Y^{\text{obs}})$: p-value = 0.80

$\Rightarrow$ In this example, the observed difference could well be due to chance

Computation of Approximate p-values

- Recall that the number of distinct values of the treatment vector is $\binom{N_c+N_t}{N_t}$

$\Rightarrow$ For instance, if $N = 100$ and $q = 0.5$ then $dim(\mathbb{W}^+) = e^{29}$

- We thus need to rely on numerical approximations to calculate the p-value
  - $\Rightarrow$ Draw an N-dimensional vector with N_c zeros and N_t ones from $\mathbb{W}^+$
  - $\Rightarrow$ Repeat this process $K - 1$ times and approximate the p-value by:

$$\widehat{p} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{1}_{T^{\text{dif},k} \geq T^{\text{dif,obs}}}$$

$\Rightarrow$ With $K > 1,000$ each assignment has a similar prob. with/out replacement

Software Implementation

- Hess (Stata Journal, 2017)

  $\Rightarrow$ `ritest` package

- Bowers et al. (Cran R project, 2024)

  $\Rightarrow$ `RItools` package

Randomization Designs

Classic Randomization Designs

$\Rightarrow$ Classic randomization designs can be characterized by restrictions on $\mathbb{W}^+$

Type of Experiment and Design	Number of Possible Assignments Cardinality of $\mathbb{W}^+$	Number of Units (N) in Sample			
		4	8	16	32
Bernoulli trial	2^N	16	256	65,536	4.2×10^9
Completely randomized experiment	$\binom{N}{N/2}$	6	70	12,870	0.6×10^9
Stratified randomized experiment	$\binom{N/2}{N/4}^2$	4	36	4,900	0.2×10^9
Paired randomized experiment	$2^{N/2}$	4	16	256	65,536

Bernoulli Trials (Coin Tossing)

- $p = e = 0.5, \mathbb{W}^+ = \{0,1\}^N$

$$P(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = 0.5^N$$

- More generally, with probability of assignment to treatment $\neq 0.5$

$$P(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = q^{N_t}(1-q)^{N_c}$$

$\Rightarrow$ No way to ensure "enough" treated and control units under each assignment

Completely Randomized Experiments

- Draw N_t units at random, such that $1 \leq N_t \leq N-1$
- $q = \frac{N_t}{N}$, and the number of possible assignments is $\binom{N}{N_t}$
- A completely randomized experiment has an assignment mechanism satisfying

$$\mathbb{W}^+ = \left\{ \mathbf{W} \in \mathbb{W} \mid \sum_{i=1}^{N} W_i = N_t \right\}$$

$\Rightarrow$ Possible issue with covariate unbalancedness after treatment assignment

Stratified Randomized Experiments

- The population of units is first partitioned into blocks or strata $B_i = B(\boldsymbol{X}_i)$
- Within each block, we conduct a completely randomized experiment
- The assignment mechanism is

$$\mathbb{W}^+ = \left\{ \mathbf{W} \in \mathbb{W} \mid \sum_{i:B_i=j}^{N} W_i = N_t(j) \right\}$$

$\Rightarrow$ More precise inference as units of same type are in different treat arms

The Benefits of Stratification (1)

- Consider a DGP with $i = 1, ..., N$ and one covariate $G_i \in \{f, m\}$
  - $\Rightarrow$ $p(G_i = f) = p = \frac{N(f)}{N}$
  - $\Rightarrow$ $p(G_i = m) = 1 - p = \frac{N(m)}{N}$
- Completely randomized design
  - $\Rightarrow$ $N_t = qN$ and $N_c = (1 - q)N$
- ATE and its sampling variance

$$\hat{\tau}^{\text{dif}} = \overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}$$
$$\mathbb{V}(\hat{\tau}^{\text{dif}}) = \frac{\sigma_t^2}{N_t} + \frac{\sigma_c^2}{N_c}$$

The Benefits of Stratification (2)

- Stratified design

  $\Rightarrow$ $N_t(f) = pqN$ and $N_c(f) = p(1-q)N$

  $\Rightarrow$ $N_t(m) = (1-p)qN$ and $N_c(m) = (1-p)(1-q)N$

- CATE (G_i)

$$\hat{\tau}^{\text{dif}}(f) = \overline{Y}_t^{\text{obs}}(f) - \overline{Y}_c^{\text{obs}}(f)$$
$$\hat{\tau}^{\text{dif}}(m) = \overline{Y}_t^{\text{obs}}(m) - \overline{Y}_c^{\text{obs}}(m)$$

- ATE and its sampling variance

$$\hat{\tau}^{\text{strat}} = p\hat{\tau}^{\text{dif}}(f) + (1-p)\hat{\tau}^{\text{dif}}(m)$$
$$\mathbb{V}(\hat{\tau}^{\text{strat}}) = \frac{p}{N}\left(\frac{\sigma_t^2(f)}{p} + \frac{\sigma_c^2(f)}{1-p}\right) + \frac{1-p}{N}\left(\frac{\sigma_t^2(m)}{p} + \frac{\sigma_c^2(m)}{1-p}\right)$$

The Benefits of Stratification (3)

- The difference in the two variances is

$$\mathbb{V}(\hat{\tau}^{\text{dif}}) - \mathbb{V}(\hat{\tau}^{\text{strat}}) = \frac{p(1-p)}{N}((\mu_c(f) - \mu_c(m))^2 + (\mu_t(f) - \mu_t(m))^2) \geq 0$$

$\Rightarrow$ The extra-variance in dif vs. strat comes from the unbalancedness of G_i

$\Rightarrow$ In principle, $\hat{V}^{\text{dif}} < \hat{V}^{\text{strat}}$ (within-stratum potential outcome variances)

Paired Randomized Experiments

- As many units as treatments within each block
- $N(j) = 2$ and $N_t(j) = 1$ for $j = 1, ..., N/2$, so that

$$\mathbb{W}^{+} = \left\{ \mathbf{W} \in \mathbb{W} \mid \sum_{i:B_i=j}^{N} W_i = 1 \right\}.$$

$\Rightarrow$ Useful design when N is small and/or J is large

Other Randomization Designs

- These experimental designs have become popular in recent years
  - Clustered randomized experiments
    - ⇒ Partial population designs
    - ⇒ Randomized saturation designs
  - Adaptive randomized experiments

Clustered Randomized Experiments

- Clusters are defined by partitioning the covariate space $G_{ig} = G(\boldsymbol{X}_i)$
- $\overline{W}_g = \sum_{i:G_{ig}=1} \frac{W_i}{N_g} \in \{0,1\}$ is the average value of W_i for units in cluster g
- The assignment mechanisms concerns groups of units (clusters)

$$\mathbb{W}^+ = \left\{ \mathbf{W} \in \mathbb{W} \mid \sum_{g=1}^{G} \overline{W}_g = G_t \right\}$$

$\Rightarrow$ Useful design when there are local interactions between units

$\Rightarrow$ Relax SUTVA within clusters, but maintain it across clusters

Partial Population Designs

- Within-cluster (non-random) program assignment $E_{ig} = \{0, 1\}$
- Write potential outcomes as $Y_i(W_g, E_{ig})$

$$\begin{aligned} ATE &= \mathbb{E}(Y_i(1,1) - Y_i(0,1)) = \\ &= \mathbb{E}(Y_i | W_g = 1, E_{ig} = 1) - \mathbb{E}(Y_i | W_g = 0, E_{ig} = 1) \\ ITE &= \mathbb{E}(Y_i(1,0) - Y_i(0,0)) = \\ &= \mathbb{E}(Y_i | W_g = 1, E_{ig} = 0) - \mathbb{E}(Y_i | W_g = 0, E_{ig} = 0) \end{aligned}$$

$\Rightarrow$ ATE and ITE depend on $S_g = \sum_{i \in g} E_{i,g}$, which is endogenous

Randomized Saturation Design

- A variant of clustered randomization where
  - ⇒ Assign each cluster to a treatment saturation, $S_g = \sum_{i \in g} W_{ig} \in [0, 1)$
  - ⇒ Assign each individual to a treatment status $W_{ig} = \{0, 1\}$ according to S_g
- Write potential outcomes as $Y_i(W_{ig}, S_g)$

$$\begin{aligned} ATE(s) &= \mathbb{E}(Y_i(1,s) - Y_i(0,0)) = \\ &= \mathbb{E}(Y_i | W_{ig} = 1, S_g = 0) - \mathbb{E}(Y_i | W_{ig} = 0, S_g = 0) + \\ &\quad + \mathbb{E}(Y_i | W_{ig} = 1, S_g = s) - \mathbb{E}(Y_i | W_{ig} = 1, S_g = 0) \\ ITE(s) &= \mathbb{E}(Y_i(0,s) - Y_i(0,0)) = \\ &= \mathbb{E}(Y_i | W_{ig} = 0, S_g = s) - \mathbb{E}(Y_i | W_{ig} = 0, S_g = 0) \end{aligned}$$

⇒ Optimal combination of clustered and stratified designs

Adaptive Randomized Experiments

- Waves $t = 1, ..., T$, sample sizes N_t
- Treatment $D \in \{1, ..., k\}$, outcomes $Y \in [0, 1]$, covariates X
- Potential outcomes Y^d, and $\theta^{dx} = E[Y_{it}^d | X_{it} = x]$
- Repeated cross-sections: $(Y_{it}^1, ..., Y_{it}^k; X_{it})$ are i.i.d. across both i and t

$\Rightarrow$ Form posterior beliefs P_t over θ, e.g. (thompson sampling):

$$p_t^{dx} = P_t\left(d = \underset{d'}{\text{argmax}}\ \theta^{d'x}\right)$$

$\Rightarrow$ Stop when $p_t^{dx} > q$ for $D = d$

Stratified randomized experiments

The Structure of Stratified Randomized Experiments

- Let J be the number of strata/blocks, and $N(j), N_c(j), N_t(j)$
- Let $S_i \in \{1, ..., J\}$ be the stratum for unit i
- Let $B_i(j) = \mathbf{1}_{S_i=j}$ be the stratum indicator for unit i
- The assignment mechanism is

$$P(\mathbf{W}|\mathbf{B}, \mathbf{Y}(0), \mathbf{Y}(1)) = \prod_{j=1}^{J} \binom{N(j)}{N_t(j)}^{-1} \text{ for } \mathbf{W} \in \mathbb{W}^+$$

$\Rightarrow$ $\mathbb{W}^+ = \{\mathbf{W} \in \mathbb{W} \mid \sum_{i=1}^{N} B_i(j) \cdot W_i = N_t(j) \text{ for } j = 1, ..., J\}$

Example: Tennessee Project Star

School/ Stratum	No. of Classes	Regular Classes ($W_i = 0$)	Small Classes ($W_i = 1$)
1	4	−0.197, 0.236	0.165, 0.321
2	4	0.117, 1.190	0.918, −0.202
3	5	−0.496, 0.225	0.341, 0.561, −0.059
4	4	−1.104, −0.956	−0.024, −0.450
5	4	−0.126, 0.106	−0.258, −0.083
6	4	−0.597, −0.495	1.151, 0.707
7	4	0.685, 0.270	0.077, 0.371
8	6	−0.934, −0.633	−0.870, −0.496, −0.444, 0.392
9	4	−0.891, −0.856	−0.568, −1.189
10	4	−0.473, −0.807	−0.727, −0.580
11	4	−0.383, 0.313	−0.533, 0.458
12	5	0.474, 0.140	1.001, 0.102, 0.484
13	4	0.205, 0.296	0.855, 0.509
14	4	0.742, 0.175	0.618, 0.978
15	4	−0.434, −0.293	−0.545, 0.234
16	4	0.355, −0.130	−0.240, −0.150
Average		−0.13	0.09
(S.D.)		(0.56)	(0.61)

Randomization Inference for Stratified Experiments

- Sharp null hypothesis

$$H_0 \; : \; Y_i(1) = Y_i(0) \; \forall i = 1, 2, ..., N.$$

- Define average observed outcomes in stratum j as

$$\overline{Y}_t^{\text{obs}}(j) = \frac{1}{N_t(j)} \sum_{i:S_i=j} W_i Y_i^{\text{obs}}$$

$$\overline{Y}_c^{\text{obs}}(j) = \frac{1}{N_c(j)} \sum_{i:S_i=j} (1 - W_i) Y_i^{\text{obs}}$$

$\Rightarrow$ Strata-level propensity score is

$$e(j) = \frac{N_t(j)}{N(j)}$$

Test Statistics

- Within-stratum test statistic

$$T^{\text{dif}}(j) = \left| \overline{Y}_t^{\text{obs}}(j) - \overline{Y}_c^{\text{obs}}(j) \right|$$

⇒ Not very informative as we are interested in treatment effects across all strata

- Linear combination of the within-stratum statistics

$$T^{\text{dif},\lambda_{RSS}} = \left| \sum_{j=1}^{J} \frac{N_j}{N} \left(\overline{Y}_t^{\text{obs}}(j) - \overline{Y}_c^{\text{obs}}(j) \right) \right|$$

⇒ Need $e(j)$ to vary across j for the test to have power over T^{dif}

Randomization Inference of the Tennessee Project Star

- $B_i(j), i = 1, ..., 68$ (class-level data)

$$H_0 \ : \ Y_i(1) = Y_i(0) \ \forall i = 1, 2, ..., 68.$$

- Total number of possible assignments is $(6^{13}) \times 10^2 \times 15$
  - $\Rightarrow$ 13 Schools with two classes in each group: $\binom{4}{2} = 6$
  - $\Rightarrow$ 2 Schools with three small classes and two regular classes: $\binom{5}{2} = 10$
  - $\Rightarrow$ 1 School with four small classes and two regular classes: $\binom{6}{2} = 15$

$\Rightarrow$ $T^{\text{dif}} = \mid \overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} \mid = 0.224$, with $p = 0.034$

$\Rightarrow$ $T^{\text{dif},\lambda_{RSS}} = \mid \sum_{j=1}^{J} \frac{N_j}{N} (\overline{Y}_t^{\text{obs}}(j) - \overline{Y}_c^{\text{obs}}(j)) \mid = 0.241$, with $p = 0.023$

Regression Analysis

$$Y_i^{\text{obs}} = \tau W_i + \sum_{j=1}^{J} \beta(j) B_i(j) + \epsilon_i$$

- Recall that $B_i(j) = \mathbf{1}_{S_i=j}$ is the stratum indicator for unit i

$\Rightarrow$ $\hat{\tau}^{\text{ols}}$ is not a consistent estimator of ATE if $\tau(j) \neq \tau(j')\ \forall j \neq j'$

$$WATE = \frac{\sum_{j=1}^{J} \omega(j)\tau(j)}{\sum_{j=1}^{J} \omega(j)}$$

- $\omega(j) = \frac{N_j}{N} \frac{N_t(j)}{N(j)} \frac{N(j)-N_t(j)}{N(j)} = q(j)e(j)(1-e(j))$

Asymptotic Inference

- The asymptotic variance of the WATE is

$$\mathbb{V}^{\text{strat}} = \frac{\sum_{i=1}^{N} \epsilon_i^2 \cdot \left(W_i - \sum_{j=1}^{J} q(j) B_i(j)\right)^2}{\left(\sum_{j=1}^{J} \omega(j)\right)^2}$$

$\Rightarrow$ Variance weights more those $\widehat{\tau}(j)$ that are more precisely estimated

$$\mathbb{V}^{j} = \frac{\sigma^2/N}{q(j)e(j)(1-e(j))}$$

Fully-interacted Model

$$Y_i^{\text{obs}} = \tau W_i \frac{B_i(j)}{N(j)/N} + \sum_{j=1}^{J} \beta(j) B_i(j) + \sum_{j=1}^{J-1} \gamma(j) W_i \left(B_i(j) - B_i(J) \frac{N(j)}{N(J)} \right) + \epsilon_i$$

$\Rightarrow$ In this case, $\hat{\tau}^{\text{ols}}$ is a consistent estimator of ATE

- With asymptotic variance

$$\mathbb{V}^{\text{strat,inter}} = \sum_{i=1}^{N} q(j)^2 \cdot \left(\frac{\sigma_c^2(j)}{q(j)(1-e(j))} + \frac{\sigma_t^2(j)}{q(j)e(j)} \right)$$

$\Rightarrow$ In general, $\mathbb{V}^{\text{strat,inter}} > \mathbb{V}^{\text{strat}}$

Regression Analysis of the Tennessee Project Star

- The point estimate of τ in the standard model is
  - $\widehat{\tau}^{\text{ols}} = 0.238$ $(\widehat{s.e.} = 0.103)$
- The point estimate of τ in the fully-interacted model is
  - $\widehat{\tau}^{\text{ols,inter}} = 0.241$ $(\widehat{s.e.} = 0.095)$

$\Rightarrow$ Limited heterogeneity in the treatment effects across strata

Sample Code: Stratified Randomization

```
set seed 123456
gen random = uniform()
egen strata=group(x1 x2)
sort strata random
by strata : gen strata_size = _N
by strata : gen strata_index = _n
gen treat = 0
replace treat = 1 if strata_index <= (strata_size/2)

areg y treat, abs(strata_var) vce(hc3)

ritest treat _b[treat]/_se[treat], reps(1000) seed(125)
strata(strata): areg y treat, abs(strata) vce(hc3)
```

Clustered Randomized Experiments

The Structure of Clustered Experiments

- Let G_{ig} be a binary indicator that unit i belongs to cluster $g = 1, ..., G$
- $N_g = \sum_{i=1}^{N} G_{ig}$, so that N_g/N is the share of cluster g in the sample
- $\overline{W}_g \in \{0, 1\}$ is the treatment assignment for all units in cluster g
- The assignment mechanism is

$$P(\mathbf{W}, \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{X}) = \binom{G}{G_t}^{-1}$$

$\Rightarrow$ $\mathbb{W}^{+} = \{\mathbf{W} \in \mathbb{W} \mid \sum_{g=1}^{G} \overline{W}_g = G_t\}$

Example: The *Progresa* Program (Again!)

- Clustered RCT during the roll-out of the program in rural areas
  - 506 villages among those eligible to receive the program
  - 320 early treatment and 186 late treatment (control)
- Individual/HH level data for both eligible and non-eligible HHs in each village
  - Approx. 30,000 program eligible children
  - About 50-100 HHs per village

Estimands

- The choice of estimand depends on the choice of the unit of analysis

⇒ Unit-level: a natural estimand is the usual ATE

$$\tau = \frac{1}{N}\sum_{i=1}^{N}(Y_i(1) - Y_i(0))$$

⇒ Cluster-level: (unweighted) average of the within-cluster average effect

$$\tau^{\text{clust}} = \frac{1}{G}\sum_{g=1}^{G}\tau_g, \quad \text{where } \tau_g = \frac{1}{N_g}\sum_{i:G_{ig}=1}^{N}(Y_i(1) - Y_i(0))$$

Unit-level Vs. Cluster-level

- Cluster-level analysis is more directly linked to the randomization framework
  - ⇒ Differently-sized clusters, such as states or towns
  - ⇒ Many units will be in the same treatment group so unit-level inference is tricky
- Unit-level incorporates individual-level covariates, which improve efficiency
  - ⇒ More homogenous clusters, such as schools or classrooms

Randomization Inference

- The usual statistic for unit-level analysis

$$T^{\text{dif}} = \mid \overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} \mid$$

- The equivalent statistic for cluster-level analysis

$$T^{\text{clust}} = \left| \frac{1}{G_t} \sum_{g:\overline{W}_g=1} \overline{Y}_g^{\text{obs}} - \frac{1}{G_c} \sum_{g:\overline{W}_g=0} \overline{Y}_g^{\text{obs}} \right|$$

Randomization Inference of *Progresa*

- Children-level analysis on school enrollment (pre-program year 1997)

$$T^{\text{dif}} = 0.0075, \;\; p\text{-value} = 0.400$$

- Children-level analysis on school enrollment (program year 1998)

$$T^{\text{dif}} = 0.0388, \;\; p\text{-value} < 0.001$$

- Village-level analysis on school enrollment (program year 1998)

$$T^{\text{clust}} = 0.0234, \;\; p\text{-value} = 0.0120$$

Regression Methods

- In unit-level analysis, we estimate the following regression

$$Y_i^{\text{obs}} = \alpha + \tau W_i + \bar{X}'\gamma + \epsilon_i$$

$\Rightarrow$ $\widehat{\tau}^{\text{ols}}$ consistently estimates ATE (with centered covariates)

- If sample has all the clusters G in population, use heterosk-robust Var

$$\mathbb{V}_{\text{rob}} = \frac{\sum_{i=1}^{N} \epsilon_i^2 \cdot \left(W_i - \overline{W}\right)^2}{\left(\sum_{i=1}^{N} (W_i - \overline{W})^2\right)^2}$$

Regression Methods

- If sample has random sub-set of clusters out of G, correct variance is:

$$\mathbb{V}_{\text{clust}} = \frac{\sum_{g=1}^{G}\left(\sum_{i:G_{ig}=1}\epsilon_i^2 \cdot \left(W_i - \overline{W}\right)^2\right)}{\left(\sum_{i=1}^{N}(W_i - \overline{W})^2\right)^2}$$

Regression Analysis: Cluster-Level

- In cluster-level analysis, consider the following regression

$$\overline{Y}_g^{\text{obs}} = \alpha + \tau \overline{W}_g + \eta_g$$

$\Rightarrow$ $\widehat{\tau}^{\text{ols}}$ consistently estimates ATE

- The sampling variance of τ^{ols} is the usual one

$$\mathbb{V} = \frac{\sum_{g=1}^{G} \eta_g^2}{\sum_{g=1}^{G} \left(\overline{W}_g - \overline{W}\right)^2} = \sigma^2 \left\{ \frac{1}{G_t} + \frac{1}{G_c} \right\}$$

Regression Analysis of *Progresa*

- Children-level analysis on school enrollment (pre-program year 1997)

$$\widehat{\tau}^{\text{ols}} = 0.0075\,(\widehat{s.e.} = 0.0091)$$

- Children-level analysis on school enrollment (program year 1998)

$$\widehat{\tau}^{\text{ols}} = 0.0388\,(\widehat{s.e.} = 0.0104)$$

- Village-level analysis on school enrollment (program year 1998)

$$\widehat{\tau}^{\text{ols}} = 0.0234\,(\widehat{s.e.} = 0.0092)$$

Sample Code: Clustered Randomization

```
* cluster-level data
set seed 123456
gen random = uniform()
sort random
gen treat = 0
replace treat = 1 if _n <= _N/2

* unit-level data
reg y treat, cluster(cluster_id)

ritest treat _b[treat]/_se[treat], reps(1000) seed(125)
cluster(cluster_id): reg y treat, cluster(cluster_id)
```

Risk sharing and seasonal migration (Meghir et al, Restud, 2022)

Migration and Informal Insurance: Evidence from a Randomized Controlled Trial and a Structural Model

$\Rightarrow$ Migration subsidies in bad times crowd in informal insurance

- A joint model of migration and informal risk sharing explains why
- Quantify the welfare effect of the migration subsidies
- Conduct counterfactual experiments to evaluate different policy levers

The Migration Experiment

- The migration subsidy treatment was randomized at the village level
  - 68 treated villages and 32 control villages
  - Random sampling of 19 eligible (poor) HHs in each village
- Rich HH-level data collected before, during, and after the intervention
  - Annual income (home, city, and total)
  - Consumption (food, and non-food)
  - Savings and transfers (received and given)

Experimental Evidence on Financial Transfers

	Treatment effect	Control mean
Willingness to help		
Community member would help you	0.030	0.85
	(0.020)	
... and you would ask for help	0.025	0.83
	(0.020)	
Community member would ask you for help	0.109***	0.57
	(0.033)	
... and you would help them	0.109***	0.53
	(0.032)	
Actual transfers		
Receive any transfer from community member	−0.024	0.57
	(0.022)	
Amount, if any transfer received (Tk)	1,821***	4808
	(678)	
Give any transfer to community member	0.036**	0.15
	(0.018)	
Amount, if any transfer given (Tk)	1,310**	2001
	(558)	

Notes: The sample includes households from the 2011 survey. Each cell is a separate regression of the effect of treatment on whether the source denoted in the row would behave as described. Each regression also controls for upazila (county). Standard errors, clustered by village, are in parentheses, and the mean of the control group is in square brackets. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Experimental Evidence on Financial Transfers

- Increased willingness/ability to provide and receive transfers btw. HHs
- This is not limited to households that were induced to migrate

$\Rightarrow$ Migration strengthened informal relationships within a village more broadly

Transmission of Income to Consumption

	Round 4				Diff in Diff	
	(1)	(2)	(3)	(4)	(5)	(6)
Log income (round 4)	0.157***	0.169***	0.130***	0.140***	0.112**	0.109**
	(0.027)	(0.028)	(0.028)	(0.029)	(0.054)	(0.046)
Treatment effect on log income	−0.073***	−0.066**	−0.072***	−0.061**	−0.077	−0.099**
	(0.027)	(0.027)	(0.027)	(0.026)	(0.061)	(0.046)
Village-round FE	X	X	X	X	X	X
Household FE					X	X
Household head controls		X		X		
Resource controls			X	X		
Includes baseline					X	X
Includes 2013						X
Observations	1857	1857	1857	1857	2166	4371
R^2	0.186	0.221	0.217	0.267	0.791	0.721

Notes: Table presents coefficients of the effect of log annual per-capita income on log annual per capita consumption and the interaction with treatment (β_0 and β_1 from equation 2). All models control for village fixed effects and all other interactions between treatment and log income as well as log income interacted with 2011 treatments. Column (2) additionally adds household head controls, column (3) adds household resource controls, and column (4) adds both household head and resource controls. Columns (5) and (6) show the result of difference-in-difference specifications, with the first coefficient shown being the interaction between log income and round 4, and the second coefficient shown being the interaction between treatment, log income, and a post-experiment indicator. Column (5) includes baseline data, and column (6) includes both baseline and 2013 data, and both include household fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Transmission of Income to Consumption

- Treat reduced the effect of HH income on consumption by over 7 pp
- Not driven by differences in permanent income (check also Appendix A.1)
- Similar effects for non-migrant HHs
- No effect on savings (and mean savings are also very small)

What is the Rationale for the Model?

- Experiment led to an improvement in the willingness and ability to share risk
- Why did this happen?

$\Rightarrow$ Migration subsidies interact with the underlying risk environment

  - Subsidies increase social return to migration (crowd-in)
  - If migration is relatively safe, migrant may not need safety net (crowd-out)

- Welfare effects of policy are heavily context dependent

A Joint Model of Risk Sharing and Migration

- Starting point is Morten (JPE, 2019)
  - Migration decisions depend on risk-sharing transfers
  - Risk sharing is constrained by limited commitment (Ligon et al, ReStud 2002)
- Add to this a migration asset (i.e. a job connection at destination)
  - ⇒ Migrants tend to return to the same employer
  - ⇒ A one-time experiment led to persistent effects on migration
- By allowing to update migration asset it provides an option value

A Joint Model of Risk Sharing and Migration

Model period	**Activities**	**State of the world**
Start of period t —	Start-of-period job contacts (A_t)	$h_t = \{s_t, A_t\}$
Before migration —	Observe village state (s_t)	
	Migration decision (j_t)	
After migration —	Exogenous creation of job contacts ($\hat{A}_t$)	$\hat{h}_t = \{h_t, j_t, q_t, \hat{A}_t(h_t, j_t)\}$
	Observe migration state (q_t)	
	Make risk-sharing transfers	
	Consume	
	Return to village	
	Exogenous separation of job contacts (A_{t+1})	
Start of period $t+1$ —	Start-of-period job contacts (A_{t+1})	$h_{t+1} = \{s_{t+1}, A_{t+1}\}$
	Observe village state (s_{t+1})	

Figure 1: Model timeline

Optimization Problem without Risk Sharing

- The before-migration value

$$\Omega^i(h) = \max_{\mathbb{I}^i} \sum_{\hat{h}} \pi_{\hat{h}|h,\mathbb{I}^i} \left[u(\hat{y}^i(\hat{h}, d^{\text{fin}})) - \mathbb{I}^i(\hat{h}) d^{\text{utility}} + \beta \sum_{h'} \pi_{h'|\hat{h}} \Omega^i(h') \right]$$

- The after-migration value

$$\hat{\Omega}^i(h) = u(\hat{y}^i(\hat{h}, d^{\text{fin}})) - \mathbb{I}^i(\hat{h}) d^{\text{utility}} + \beta \sum_{h'} \pi_{h'|\hat{h}} \Omega^i(h')$$

$\Rightarrow$ $\Omega^i(h)$ and $\hat{\Omega}^i(h)$ determine the credible threat points under risk sharing

Optimization Problem with Risk Sharing

- The after-migration problem is the following

$$\widehat{V}(\hat{h}, \widehat{U}(\hat{h})) = \max_{\tau(\hat{h}, d^{\text{fin}}), U(h')} u(\hat{y}^2(\hat{h}, d^{\text{fin}}) + \tau(\hat{h}, d^{\text{fin}})) - \mathbb{I}^i(\hat{h}) d^{\text{utility}} + \beta \sum_{h'} \pi_{h'|\hat{h}} V(h', U(h'))$$

- Subject to a promise-keeping constraint for HH1

$$(\widehat{\lambda}_{\hat{h}}): \; u(\hat{y}^1(\hat{h}, d^{\text{fin}}) - \tau(\hat{h}, d^{\text{fin}})) - \mathbb{I}^i(\hat{h}) d^{\text{utility}} + \beta \sum_{h'} \pi_{h'|\hat{h}} V(h', U(h')) \geq \widehat{U}(\hat{h})$$

- And incentive compatibility constraints (with LMs $\phi^1_{h',\hat{h}}$ and $\phi^2_{h',\hat{h}}$)

Optimization Problem with Risk Sharing

- FOCs yield (see Appendix A.2)

$$\frac{u_1(c^2(\hat{h}))}{u_1(c^1(\hat{h}))} = \hat{\lambda}_{\hat{h}}$$

$$V_2(h', U(h')) = \widehat{\lambda}_{\hat{h}} \frac{(1+\phi^1_{h',\hat{h}})}{(1+\phi^2_{h',\hat{h}})}$$

- The envelope condition yields

$$\widehat{V}_2(\hat{h}, \widehat{U}(\hat{h})) = \hat{\lambda}_{\hat{h}}$$

Optimization Problem with Risk Sharing

- The before-migration problem is the following:

$$V(h, U(h)) = \max_{j} \left\{ \max_{\widehat{U}(\hat{h})} \left[\sum_{\hat{h}} \pi_{\hat{h}|h,j} \hat{V}(\widehat{h}, \widehat{U}(\hat{h})) \right] \right\}$$

  - Subject to promise-keeping constraint with LM λ_j
  - And incentive compatibility constraints with LMs $\alpha^1_{\hat{h}}$ and $\alpha^2_{\hat{h}}$

$\Rightarrow$ FOC and envelope condition yield

$$\widehat{V}_2(\hat{h}, U(\hat{h})) = -\lambda_j \frac{(1+\alpha^1_{\hat{h}})}{(1+\alpha^2_{\hat{h}})} \ \forall \hat{h}$$

$$V_2(h, U(h)) = -\lambda_j \ \forall j$$

Updating Rules

- Pareto weight follows a simple updating rule

$$\frac{u_1(c^2(\hat{h}_{t+1}))/u_1(c^1(\hat{h}_{t+1}))}{u_1(c^2(\hat{h}_t))/u_1(c^1(\hat{h}_t))} = \frac{\hat{\lambda}_{t+1}}{\hat{\lambda}_t} = \frac{(1+\phi^1_{h_{t+1},\hat{h}_t})(1+\alpha^1_{h_{t+1}})}{(1+\phi^2_{h_{t+1},\hat{h}_t})(1+\alpha^2_{h_{t+1}})}$$

$\Rightarrow$ If neither HHs is constrained, then the growth rate of relative MU is zero

$\Rightarrow$ Otherwise, the change in MU adjusts towards HH with binding constraints

Solving the Model

- The model can be extended from 2 HHs to N HHs
  - An aggregated "rest of the village" HH
  - $\Rightarrow$ Simulate N households who each follow the policy rule from the two-HHs case
  - One additional parameter to satisfy the economy-wide budget constraint
- Computational algorithm is quite involved (check Appendix A.3)
  - Draws from Morten (2019) + migration asset and temp. experiment shock
  - $\Rightarrow$ Solution by value function iteration with before- and after-migration grid

From the Experiment to the Model

$\Rightarrow$ The experiment changes the financial cost of migrating: d_t^{fin}

$\Rightarrow$ Additional utility cost of migrating: $\Delta d_t^{\text{utility}}$

$\Rightarrow$ Changes the value of autarky for HHs

Identification (focusing on the experimental variation)

- Parameters of the migration asset (πs) identified off migration transitions
  - $\Rightarrow$ Re-migration rate following the first migration episode
- Migration parameters (opportunity cost, utility cost and subsidy)
  - $\Rightarrow$ Treatment effect on migration rates, both during and after the experiment
- Other preference parameters (risk aversion and discount factor)
  - $\Rightarrow$ Consumption regression on simulated control and treatment group

Parameter Estimates

Preferences	
CRRA parameter	1.88
	(0.037)
Opportunity cost of migration	0.15
	(0.088)
Utility cost of migrating	0.075
	(0.0051)
Utility subsidy	0.075
	(0.025)
Decay rate of utility subsidy	0.15
	(0.83)
Income processes	
Mean home income	2.23
	(0.23)
Std. home income	0.58
	(0.0040)
Mean city income with contact	0.38
	(0.17)
Std. city income with contact	0.78
	(0.074)
Dynamics	
Prob. get contact	0.79
	(0.28)
Prob. lose contact if migrate	0.44
	(0.85)
Prob. lose contact if don't migrate	0.66
	(5.62)
Model criterion	1.715

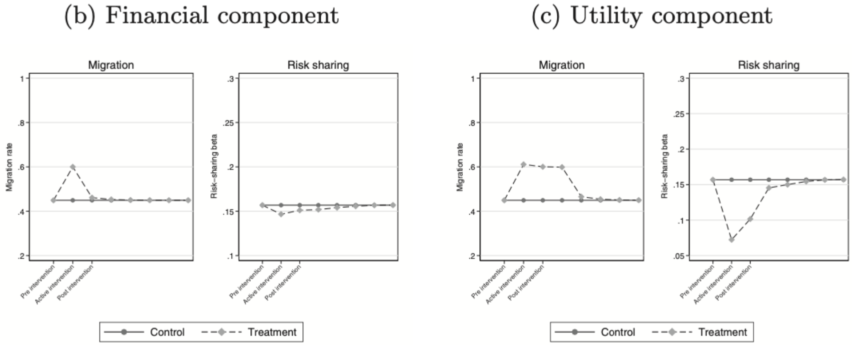
Notes: The table shows parameter estimates and standard errors. The parameter estimates arise from estimating the model by simulated method of moments. The analytical standard errors are computed by numerical differentiation. The mean level of utility in control villages is 3.23.

Model Fit

	Data	Model
Risk sharing (control)	0.16	0.16
Risk sharing (treatment effect)	−0.073	−0.055
Mean migration rate	0.39	0.45
Mig. treatment effect (during RCT)	0.22	0.35
Mig. treatment effect (after RCT)	0.094	0.15
Migrate neither during/after RCT (control)	0.49	0.32
Migrate during and after RCT (control)	0.23	0.22
Migrate neither during/after RCT (treatment effect)	−0.17	−0.16
Migrate during and after RCT (treatment effect)	0.15	0.14
Mean log home income (migrant)	1.80	1.57
Std. log home income (migrant)	0.67	0.29
Mean log home income (nonmigrant)	2.13	2.63
Std. log home income (nonmigrant)	0.56	0.33
Log std. mig. income (migrant)	0.27	0.23
Log mean mig. income (nonmigrant prior pd.)	0.60	0.39
Log mean mig. income (migrant prior pd.)	0.73	0.51
Model criterion		1.72

Notes: The table shows the targeted moments in the data (column (1)) and in the estimated model (column (2)).

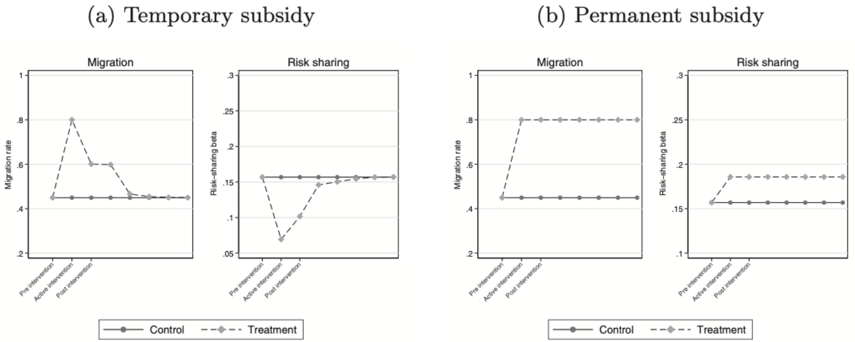
Simulating the Experiment Inside the Model



$\Rightarrow$ $\uparrow$ welfare $\approx$ to a permanent 12.9% increase in consumption (net of subsidy)

$\Rightarrow$ Welfare gains are three times higher due spillover effect through risk sharing

Counterfactuals



⇒ Permanent subsidy leads to less risk sharing (↑ outside option + ↓ migr. risk)

⇒ Difficult to extrapolate from RCTs to alternative longer-term policies

RCT ⇔ Economic Model

- RCT+Model to explore the trade-offs of migr. subsidies in village economy
- Structural model adds a dynamic component (migration state-dependent)
- RCT as source of variation + accounts for utility aspects of the subsidy

The production of human capital in the early years (Attanasio et al, AER 2020)

Estimating the Production Function for Human Capital: Results from a Randomized Controlled Trial in Colombia

- Human capital formation is a complex process
  - Human capital is multi-dimensional (cognitive, non-cognitive, health,...)
  - Skill formation is a dynamic process
  - The components of human capital interact both within and across periods
  - Both genetic and environmental factors are important inputs

$\Rightarrow$ The early years is a particularly salient period for policy

  - Human capital is malleable (and vulnerable)
  - Dynamic complementarities (skills beget skills)
  - There is no equity-efficiency trade-off for early investment

The Intervention and Its Evaluation

- Adapt a prominent ECD program in Jamaica to the Colombian context
  - Children 12-24 months living in families targeted by a cash transfer program
  - Weekly home visits of one hour from trained women from the community
  - Micronutrients supplementation
  - Cheap and scalable intervention

⇒ Clustered RCT design, stratified by region

- 1,429 children in 96 municipalities in central Colombia
- Municipalities were randomized into 4 groups (cross-cutting or factorial design)

⇒ Control

⇒ Stimulation (home visits)

⇒ Supplementation (micronutrients)

⇒ Stimulation+Supplementation

Rich Data Collected on Children and Parents

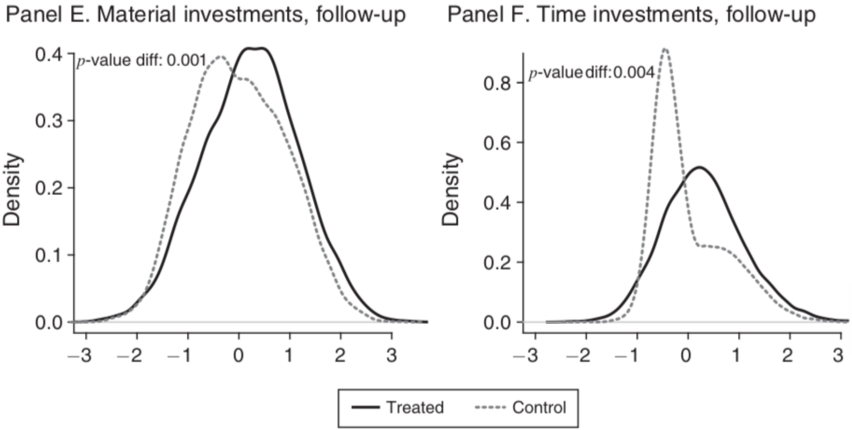
- A battery of psychometrics on children development (baseline and follow-up)
  - ⇒ Motor and cognitive skills: Bayley test
  - ⇒ Socio-emotional skills: Bates temperament test
  - ⇒ Language development: MacArthur-Bates test
  - ⇒ Height, weight, hemoglobin, and morbidity
  - ⇒ Food intakes
  - ⇒ Child care arrangements and time use
- Mothers and families
- Home visitor
- Qualitative instruments

Experimental Results

⇒ Stimulation improved cognitive scores by 0.25σ+receptive language by 0.18σ

- Also positive (noisy) effects on socio-emotional skills
- Micronutrient supplementation had no significant effect on any outcome
- No effect of the intervention on height, weight, or hemoglobin levels

A First Hint at the Mechanisms: Parental Investment



What is the Rationale for the Model?

- Intervention changes human capital through three different channels
  - ⇒ Total factor productivity
  - ⇒ The productivity of the individual inputs
  - ⇒ The distribution of inputs (and in particular parental investment)
- Jointly estimate the distribution of inputs and the production function
- Unpack the role of these different hypotheses

Measurement Equations

$$m^{\theta}_{j,d} = \mu^{k}_{j,d} + \alpha^{k}_{j} \ln \theta^{k}_{d} + \epsilon^{k}_{j}$$

- Treated and control children are allowed to have different factor distributions
- Latent factors are distributed as a mixture of two normal distributions
- Measurement errors terms are jointly normal
- Likelihood function for each group writes as

$$L_i(m) = \int_{\theta} f(m_i|\theta)[\tau\phi_A(\theta) + (1-\tau)\phi_b(\theta)]d\theta$$

$\Rightarrow$ $f(m_i|\theta)$ is the density of the measurement system

Signal-to-noise Ratios: $s_{j,d}^{ln\theta^k} = \frac{(\alpha_j^k)^2 var(ln\theta^k)}{(\alpha_j^k)^2 var(ln\theta^k)+var(\epsilon_j^k)}$

Latent factor	Measurement	Survey	Control	Treatment
Child's cognitive skills at FU (θ_{t+1}^C)	Bayley: Cognitive	FU	0.78	0.79
	Bayley: Receptive language	FU	0.75	0.76
	Bayley: Expressive language	FU	0.78	0.79
	Bayley: Fine motor	FU	0.59	0.61
	MacArthur: Words the child can say	FU	0.64	0.65
	MacArthur: Complex phrases the child can say	FU	0.52	0.54
Child's cognitive skills at BA (θ_t^C)	Bayley: Cognitive	BA	0.70	0.70
	Bayley: Receptive language	BA	0.73	0.72
	Bayley: Expressive language	BA	0.75	0.74
	Bayley: Fine motor	BA	0.60	0.59
	MacArthur: Words the child can say	BA	0.45	0.44
Child's socio-emotional skills at FU (θ_{t+1}^S)	ICQ: Difficult (-)	FU	0.74	0.71
	ICQ: Unsociable (-)	FU	0.33	0.30
	ICQ: Unstoppable (-)	FU	0.59	0.55
	ECBQ: Inhibitory control	FU	0.73	0.69
	ECBQ: Attentional focusing	FU	0.27	0.24
Child's socio-emotional skills at BA (θ_t^S)	ICQ: Difficult (-)	BA	0.68	0.71
	ICQ: Unsociable (-)	BA	0.28	0.31
	ICQ: Unadaptable (-)	BA	0.35	0.38
	ICQ: Unstoppable (-)	BA	0.22	0.25

Treatment Impacts on Measures and Factors

	Treatment effect		
	Point estimate	SE	Sample size
Panel A. Child's cognitive skills at follow-up			
Bayley: Cognitive	0.250	(0.063)	1,264
Bayley: Receptive language	0.175	(0.063)	1,264
Bayley: Expressive language	0.032	(0.062)	1,263
Bayley: Fine motor	0.072	(0.060)	1,262
MacArthur: Words the child can say	0.092	(0.064)	1,322
MacArthur: Complex phrases the child can say	0.058	(0.055)	1,322
Cognitive factor	0.115	(0.051)	
Panel B. Child's socio-emotional skills at follow-up			
ICQ: Difficult (-)	0.074	(0.045)	1,326
ICQ: Unsociable (-)	0.041	(0.054)	1,326
ICQ: Unstoppable (-)	0.032	(0.054)	1,326
ECBQ: Inhibitory control	−0.003	(0.058)	1,323
ECBQ: Attentional focusing	0.069	(0.049)	1,323
Socio-emotional factor	0.087	(0.044)	
Panel C. Material investments at follow-up			
FCI: Number of types of play materials	0.215	(0.064)	1,326
FCI: Number of coloring and drawing books	−0.133	(0.056)	1,326
FCI: Number of toys to learn movement	−0.048	(0.065)	1,326
FCI: Number of toys to learn shapes	0.416	(0.088)	1,326
FCI: Number of shop-bought toys	0.024	(0.061)	1,326
Material investment factor	0.227	(0.069)	
Panel D. Time investments at follow-up			
FCI: Number of types of play activities in last 3 days	0.277	(0.050)	1,326
FCI: Number of times told a story to child in last 3 days	0.138	(0.060)	1,326
FCI: Number of times read to child in last 3 days	0.362	(0.062)	1,326
FCI: Number of times played with toys in last 3 days	0.175	(0.060)	1,326
FCI: Number of times named things to child in last 3 days	0.137	(0.048)	1,326
Time investment factor	0.302	(0.068)	

The Production Function for Human Capital

- Cobb-Douglas specification for the technology of skill formation

$$\ln(\theta^k_{i,d,t+1}) = A^k_d + \gamma^k_{1,d}\ln(\theta^C_{i,t}) + \gamma^k_{2,d}\ln(\theta^S_{i,t}) + \gamma^k_{3,d}\ln(P^C_{i,t}) + \gamma^k_{4,d}\ln(P^S_{i,t})$$
$$+ \gamma^k_{5,d}\ln(I^M_{i,t+1}) + \gamma^k_{6,d}\ln(I^T_{i,t+1}) + \gamma^k_{7,d}\ln(n_{i,t}) + \eta^k_{i,t+1}$$

- A^k_d captures the direct effect of the home-visitor during her weekly visit

$\Rightarrow$ All parameters are skill-specific and depend on treatment status

Parental Investments

- Same log-linear specification for investment equation ($\tau = \{M, T\}$)

$$\ln(I^{\tau}_{i,t+1}) = \lambda^{\tau}_{0,d} + \lambda^{\tau}_{1,d}\ln(\theta^{C}_{i,t}) + \lambda^{\tau}_{2,d}\ln(\theta^{S}_{i,t}) + \lambda^{\tau}_{3,d}\ln(P^{C}_{i}) + \lambda^{\tau}_{4,d}\ln(P^{S}_{i}) + \lambda^{\tau}_{5,d}\ln(n_{i,t}) + \lambda^{\tau}_{6,d}Z_{i,t} + u_{i,t+1}$$

- Z_t are instrument
- All parameters are allowed to vary with the treatment
- Control function to account for correlation between shocks and investments

$\Rightarrow$ Include residuals $\hat{u}_{i,t+1}$ as additional regressor in the production function

Evidence Supporting the Exclusion Restrictions

TABLE 3—BALANCE TEST FOR THE INSTRUMENTAL VARIABLES

	log toy price	log food price	Conflict
Mother's cognitive skill	0.021	0.007	−0.004
	(0.009)	(0.007)	(0.003)
Mother's socio-emotional skill	−0.006	−0.014	0.005
	(0.008)	(0.006)	(0.005)
Mother is married	0.013	0.030	0.017
	(0.016)	(0.012)	(0.008)
Wealth index	−0.003	0.002	0.002
	(0.007)	(0.006)	(0.004)
Terrorism	0.015	−0.012	0.005
	(0.014)	(0.013)	(0.009)
Constant	8.025	8.062	0.049
	(0.028)	(0.017)	(0.007)
Observations	1,010	1,023	1,023
R^2	0.011	0.021	0.017
F-statistic	1.656	2.147	1.438
F-test p-value	0.154	0.0671	0.219

Notes: All right-hand-side variables measured at baseline. Asymptotic standard errors in parentheses allowing for clustering at the municipality level. *Terrorism* is the number of terrorist attacks between conception of child and baseline. *Conflict* is the number of conflicts against civil population divided by the municipality population (in thousands) when the mother herself was a child.

Estimation in Three Steps

⇒ Estimate the cov. of the factors based on cov. of the observed measures

⇒ Use these estimates to predict factor scores: $\hat{\theta}^k_{i,t}, \hat{\theta}^k_{i,t}$

⇒ Use predicted scores as data to estimate the parental inv. and prod. function

- Compute conf. intervals for test statistics using the block bootstrap

Investment Functions

TABLE 5—ESTIMATES OF THE MATERIAL AND TIME INVESTMENT EQUATIONS

	Instruments: prices and conflict		Instruments: prices only
	Material investment	Time investment	Material investment
Intercept	−0.015 [−0.114, 0.078]	0.001 [−0.089, 0.089]	−0.013 [−0.11, 0.078]
Treatment	0.209 [0.038, 0.365]	0.318 [0.155, 0.48]	0.204 [0.037, 0.362]
log child's cognitive skill (t)	0.130 [0.016, 0.246]	0.068 [−0.044, 0.18]	0.132 [0.017, 0.25]
log child's socio-emotional skill (t)	−0.028 [−0.133, 0.087]	0.027 [−0.083, 0.145]	−0.030 [−0.131, 0.088]
log mother's cognitive skill	0.748 [0.582, 0.939]	0.349 [0.162, 0.498]	0.750 [0.583, 0.943]
log mother's socio-emotional skill	0.069 [−0.008, 0.139]	0.022 [−0.06, 0.108]	0.068 [−0.008, 0.139]
log number of children	−0.129 [−0.18, −0.077]	−0.128 [−0.186, −0.072]	−0.128 [−0.18, −0.078]
log toy price	−0.096 [−0.168, −0.027]	−0.020 [−0.085, 0.037]	−0.094 [−0.163, −0.026]
log food price	0.091 [0.006, 0.178]	0.042 [−0.026, 0.121]	0.091 [0.006, 0.178]
Maternal childhood exposure to conflict	−0.009 [−0.08, 0.063]	−0.089 [−0.139, −0.032]	

Production Function of Cognitive Skills

TABLE 6—ESTIMATES OF THE PRODUCTION FUNCTION FOR COGNITIVE SKILLS

Instruments:	OLS	IV			
	(1)	Prices, conflict (2)	Prices, conflict, treatment (3)	Prices (4)	Prices and treatment (5)
Intercept	−0.018 [−0.094, 0.053]	−0.019 [−0.111, 0.079]	0.007 [−0.089, 0.104]	0.003 [−0.091, 0.072]	−0.009 [−0.083, 0.058]
Treatment	0.083 [−0.025, 0.192]	0.049 [−0.12, 0.391]		−0.028 [−0.186, 0.156]	
log child's cognitive skill (t)	0.675 [0.589, 0.77]	0.648 [0.544, 0.795]	0.638 [0.522, 0.747]	0.626 [0.525, 0.746]	0.631 [0.533, 0.739]
log child's socio-emotional skill (t)	0.001 [−0.091, 0.087]	0.012 [−0.098, 0.143]	0.019 [−0.094, 0.14]	0.015 [−0.084, 0.126]	0.02 [−0.079, 0.127]
log mother's cognitive skill	0.213 [0.089, 0.35]	−0.075 [−0.456, 0.5]	−0.173 [−0.538, 0.201]	−0.102 [−0.495, 0.291]	−0.094 [−0.45, 0.21]
log mother's socio-emotional skill	0.103 [0.031, 0.173]	0.084 [−0.03, 0.163]	0.063 [−0.035, 0.151]	0.06 [−0.019, 0.152]	0.074 [−0.026, 0.155]
log number of children	0.042 [−0.01, 0.092]	0.085 [−0.07, 0.154]	0.084 [0.011, 0.163]	0.089 [0.002, 0.176]	0.086 [0.023, 0.164]
log material investment	0.088 [0.016, 0.157]	0.594 [0.025, 1.179]	0.784 [0.204, 1.383]	0.542 [0.041, 0.996]	0.516 [0.195, 0.946]
log time investment	0.038 [−0.051, 0.129]	−0.171 [−1.198, 0.312]	−0.311 [−0.985, 0.217]		
Goodness-of-fit: Gap in output between treated and control					
Measured in the data	0.115	0.115	0.115	0.115	0.115
Predicted by the model	0.114	0.132	0.084	0.095	0.117

Production Function of Socio-emotional Skills

TABLE 7—ESTIMATES OF THE PRODUCTION FUNCTION FOR SOCIO-EMOTIONAL SKILLS

Instruments:	OLS	IV			
		Prices, conflict	Prices, conflict	Prices, conflict, treatment	Prices, conflict, treatment
	(1)	(2)	(3)	(4)	(5)
Intercept	−0.009 [−0.08, 0.063]	−0.02 [−0.087, 0.064]	−0.006 [−0.084, 0.066]	−0.02 [−0.096, 0.058]	−0.022 [−0.089, 0.051]
Treatment	−0.011 [−0.124, 0.093]	−0.071 [−0.318, 0.166]	−0.116 [−0.31, 0.115]		
log child's cognitive skill (t)	0.106 [0.018, 0.192]	0.074 [−0.019, 0.21]	0.09 [−0.017, 0.194]	0.094 [0.003, 0.22]	0.099 [0.002, 0.193]
log child's socio-emotional skill (t)	0.522 [0.403, 0.659]	0.513 [0.374, 0.672]	0.499 [0.389, 0.663]	0.516 [0.387, 0.669]	0.512 [0.403, 0.656]
log mother's cognitive skill	−0.077 [−0.217, 0.049]	−0.146 [−0.443, 0.297]	−0.15 [−0.349, 0.084]	−0.018 [−0.315, 0.328]	−0.083 [−0.231, 0.078]
log mother's socio-emotional skill	0.037 [−0.058, 0.119]	0.035 [−0.062, 0.141]	0.048 [−0.053, 0.135]	0.049 [−0.054, 0.134]	0.043 [−0.042, 0.126]
log of number of children	0.099 [0.047, 0.153]	0.127 [0.017, 0.236]	0.133 [0.034, 0.223]	0.101 [0.028, 0.168]	0.101 [0.036, 0.163]
log material investment	0.154 [0.06, 0.256]	0.015 [−0.621, 0.428]		−0.144 [−0.688, 0.34]	
log time investment	0.109 [−0.006, 0.213]	0.487 [−0.177, 1.258]	0.549 [−0.022, 1.147]	0.448 [−0.125, 1.057]	0.324 [0.025, 0.691]
Goodness-of-fit: Gap in output between treated and control					
Measured in the data	0.087	0.087	0.087	0.087	0.087
Predicted by the model	0.057	0.080	0.050	0.103	0.098

Interpreting the Impact of the Intervention

- Cognitive (Socio-emotional) development is mainly determined by
  - Initial cognition (measured at ages 1-2)
  - Stimulation provided by older siblings
  - Material (time) investments provided by parents

$\Rightarrow$ The main channel of the intervention is a shift in parental investments

Complementarities between Inputs

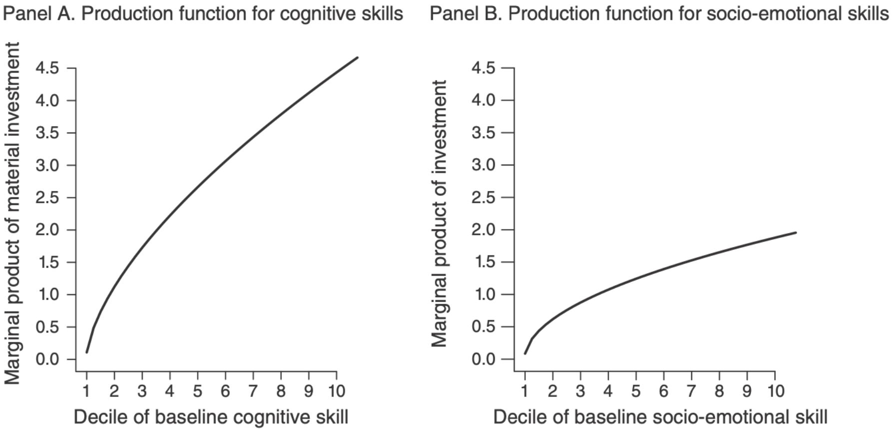


FIGURE 2. COMPLEMENTARITY BETWEEN INVESTMENTS AND BASELINE SKILLS

Implications for longer-term outcomes

- Parental crowding-in suggests treatment will diverge in the medium/long run
- But also possible fade-out due to mean reversion
- Follow-up data (2 years after) suggests partial fade out

$\Rightarrow$ How can we permanently improve parental practices ?

RCT $\Leftrightarrow$ Economic Model

- Model to interpret how the intervention affected child development

$\Rightarrow$ Sometimes the experimental variation is what you want to explain

- Need additional sources of variation beyond the experiment