Causal Inference on the Effectiveness of Outbound Calling Strategies:

Evidence from China Unicom

LIU, Liangjie

MSc in Data-Driven Modeling
The Hong Kong University of Science and Technology

May 20, 2025

Why This Study Matters

- Telecom companies often call users to promote products.
- But do these calls really make users subscribe?
- **Correlation is not causation**—just calling more doesn't prove it works.
- This study: Uses **causal inference** to estimate the true effect of outbound calls, and provides **data-driven guidance** for smarter marketing.

What We Asked — and How We Answered It

- **Q1: Does calling users more lead to more subscriptions?**
- **Q2: Does the length of the call matter?**
  - $\rightarrow$ Use a **causal framework** (DAG + PSM) to compare subscription rates
  - $\rightarrow$ Estimate **marginal effects** by call duration (dose-response curves)
- **Q3: Is the key just dialing, or actually reaching the user?**
  - $\rightarrow$ Use **mediation analysis** to separate direct vs. indirect effects via connection

Data Overview and Raw Observations

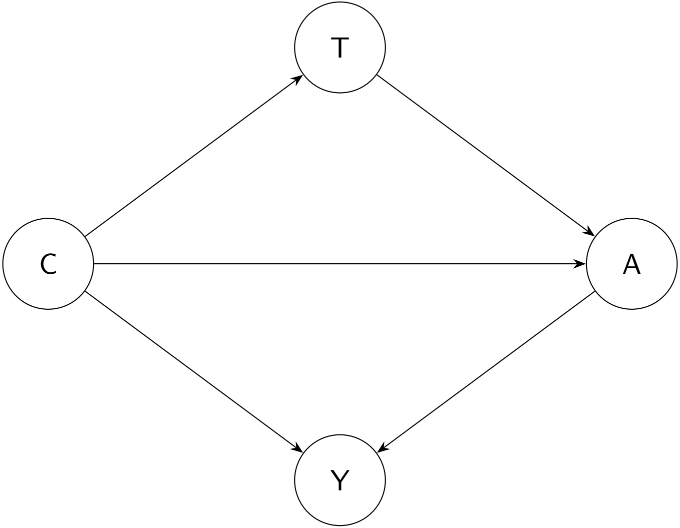
- **Data Source:** China Unicom (Dec 2024 – Jan 2025)
- **Sample:** 300,000 users, 870,000 calls, 80,000 successful contacts
- **Features include:**
  - **User info:** age, plan type, usage behavior
  - **Call logs:** number of calls, success, duration
  - **Outcome:** whether they subscribed (yes/no)

Subscription Rates by Call Activity (Unadjusted)

Dimension	Treated (>0)	Untreated (=0)	Difference
Call Attempt	8.65%	2.23%	+6.42 pp
Call Connected	17.40%	2.39%	+15.01 pp
Call Duration > 0	11.86%	4.77%	+7.09 pp

Note: These raw gaps are not causal—they may reflect underlying user differences.

Understanding the Causal Links (DAG)



User characteristics (C) are confounders, which influence treatment (T), contact success (A), and subscription (Y).

What Is ATT—and Why It Matters

- We want to know the **causal effect** of calls on subscription.
- ATT(Average Treatment Effect on the Treated) to **quantify** the effect among users who actually received calls.

- **Formula:**

$$\text{ATT} = \mathbb{E}[Y(1) - Y(0) \mid T = 1]$$

  - $Y(1)$: outcome if called
  - $Y(0)$: outcome if not called

- But we can't observe $Y(0)$ for called users ($T = 1$), so we need a way to estimate $Y(0)$.
- The challenge: **called users may differ from uncalled users**. (Users who are called may already be more likely to subscribe)

How Do We Estimate ATT? Use PSM

- We need to construct a fair comparison group $\rightarrow$ **Propensity Score Matching (PSM)**.
- Steps:
  - Estimate each user's probability of being called (the "**propensity score**").
  - Match each treated user with an untreated one with a similar score.
- This gives us a balanced sample where the only difference is: one got called, the other didn't. They are **similar**.
- But to estimate these propensity scores, we must decide:
  Which features should go into the model?
  In other words, **how do we identify the confounders?**

What Could Confuse the Results?

- We used two XGBoost models:
  - One to predict who gets called.
  - One to predict who subscribes.
- Then we used **SHAP** (SHapley Additive exPlanations) to understand which features were most important in each model.
- SHAP shows how much each feature contributes to a prediction.
- We took the top features from both models.
- **Overlapping features** = key confounders

We used SHAP to find features that affect both:

- Whether a user receives a successful call (treatment)
- Whether the user subscribes (outcome)

Top Confounders Identified by SHAP

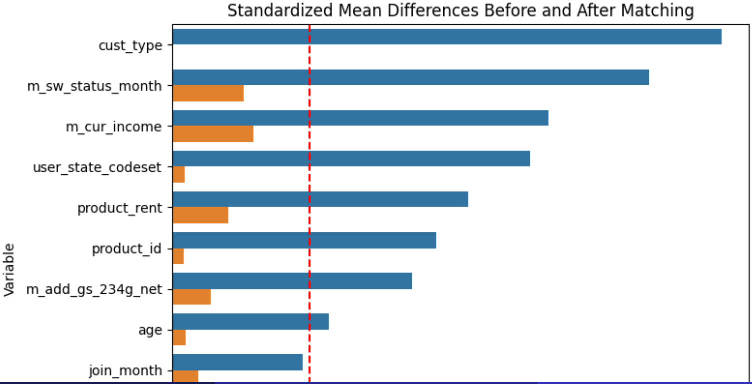
12 Shared Confounders (examples below):

- `join_month` – How long they've subscribed
- `product_id` – Current plan ID
- `product_rent` – Monthly fee
- `m_cur_income` – Current spending level
- `m_add_duration_01` – Call duration (last month)
- `m_num_bill_duration` – Total billed minutes
- `m_add_call_times` – Outbound call count
- ... (see Appendix for full list)

These variables were selected by their importance in both models (treatment and outcome).

Was the Matching Successful? (PSM Balance Check)

- Then we used **logistic regression** to estimate propensity scores and did **1:1 matching**.
- To evaluate matching quality, we checked **Standardized Mean Differences (SMD)** across 12 key confounders.
- **Result:** After matching, all SMD values dropped below 0.1 — suggesting strong covariate balance.



ATT Estimates with Comparison

Review:

- ATT tells us the **causal effect for users who actually received treatment (calls)**.

Treatment	**Unmatched Diff**	**ATT (Matched)**	**95% CI**
Call attempt	+6.42%	+6.64%	[6.49%, 6.79%]
Call connected	+15.01%	+15.24%	[14.91%, 15.56%]
Call duration > 0	+7.09%	+8.97%	[8.58%, 9.38%]

Causal effects exist!
ATT(connected)=15.24% > ATT(attempt)=6.64%
Connection is more effective than attempt!

Uncertainty: All ATT estimates are based on 1:1 matching with 95% CIs computed via **non-parametric bootstrapping**.

Insight 1: Wrong Users Are Being Targeted

- Our results show: **ATT (after PSM)** is consistently higher than raw differences.
- This suggests the current outreach strategy is **not reaching the most responsive users**.
- Likely causes:
  - Targeting based on static traits (e.g. past behavior)
  - Not accounting for **causal responsiveness**
- **Recommendation:**
  Use **uplift modeling** to:
  - Estimate individual-level treatment effects
  - Prioritize users most likely to respond causally
  - Improve ROI and reduce wasted calls

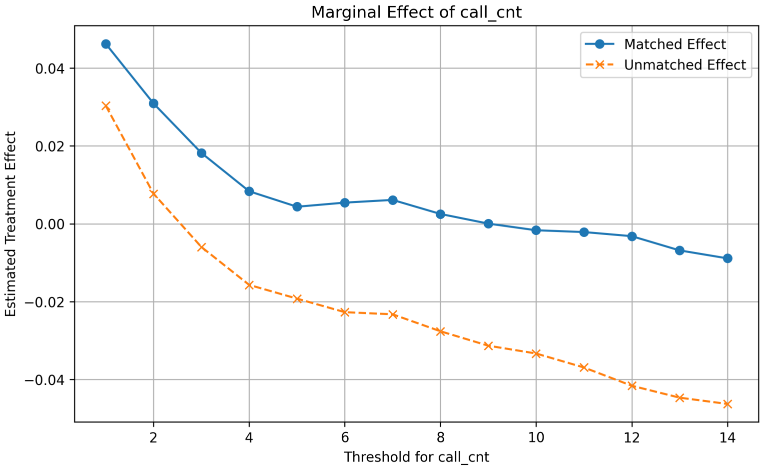
How Did We Estimate Marginal Effects?

Let's continue with Q2 and Q3.
Q2:

- We wanted to know: **Does more calling lead to better results?**
- So we used a **dose–response analysis**:
  - Define several thresholds: 1, 2, 3, ..., 8+ calls
  - For each level, estimate ATT using matched samples
- We did this for:
  1. Total call attempts
  2. Successful call count
  3. Total call duration (in seconds)
- This shows how the **effectiveness changes as intensity increases**.

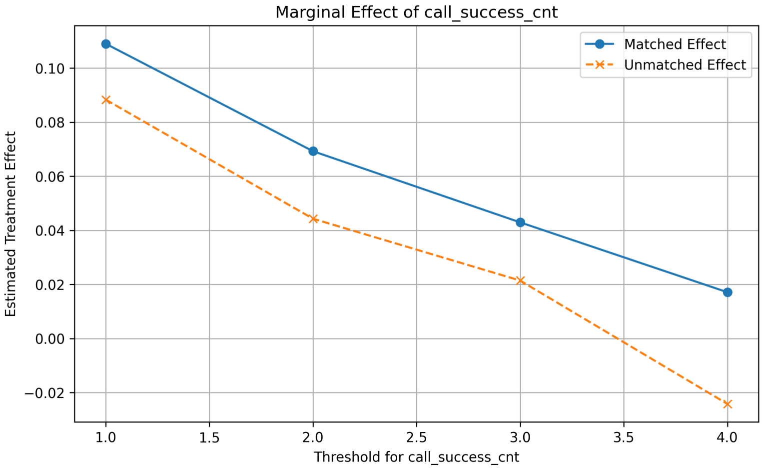
Marginal Effect: Call Attempts

- The **first few call attempts** bring the largest improvement.
- After the first few, the effect **drops**.
- Suggests **diminishing returns**—don't over-call users.



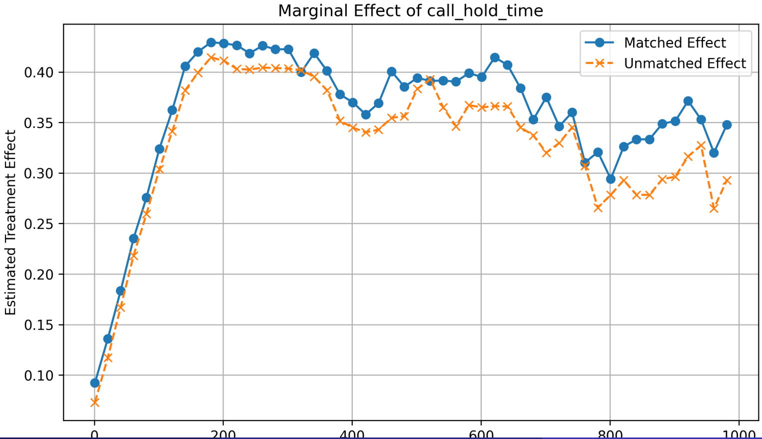
Marginal Effect: Successful Calls

- Like total attempts, the first successful call delivers the most value.
- Adding more successful calls doesn't help much after 1–2.
- Reinforces that **quality beats persistence**.



Marginal Effect: Call Duration

- Conversion rate **increases with call duration**—especially before **200 seconds**.
- After that, the effect **fluctuates with no clear gain**.
- **Takeaway:** Longer calls help—**but only up to a point**.



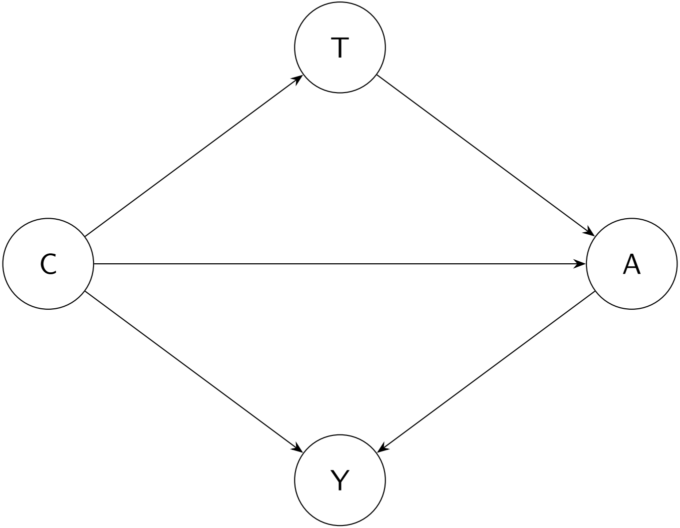
How Did We Analyze Mediation?

Q3:

- We wanted to know: **Is the key just dialing, or actually reaching the user?**
- We treat **call connection (A)** as a **mediator** between:
  - Treatment: whether a user was called (T)
  - Outcome: whether they subscribed (Y)
- We decompose the total effect of $T \rightarrow Y$ into:
  - **Indirect effect:** $T \rightarrow A \rightarrow Y$
  - **Direct effect:** $T \rightarrow Y$ (not through A)
- We used a simple **product-of-coefficients approach**:

$$\text{Indirect Effect} \approx [P(A = 1 \mid T = 1) - P(A = 1 \mid T = 0)] \times \text{ATT}_{A \rightarrow Y}$$

Review the Causal Links (DAG)



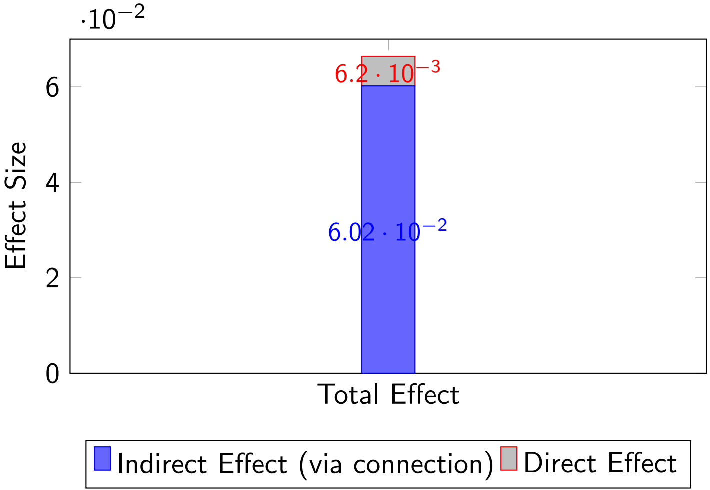
User characteristics (C) influence treatment (T), contact success (A), and subscription (Y).

Why Do Calls Work? (Mediation Results)

Quantity	Value
Connection Rate ($P(A = 1 \mid T = 1)$)	0.4075
Conversion if Connected ($P(Y = 1 \mid A = 1)$)	17.40%
Conversion if Not Connected ($P(Y = 1 \mid A = 0)$)	2.64%
Pickup Effect (Δ_{pickup})	14.76 pp
Indirect Effect ($T \to A \to Y$)	**+6.02 pp**
Direct Effect ($T \to Y$ without A)	**+0.62 pp**
Total Effect (ATT)	**+6.64 pp**

Over 90% of the total effect is explained by call connection!

Mediation Breakdown (Visual)



Over 90% of the causal effect is due to successful connection (A)!

Insight 2: From Marginal and Mediator Analysis

- First contact gives the biggest lift.
- More attempts $\rightarrow$ diminishing return.
- Longer calls help—**up to $\sim$200 seconds**.
- Causal impact comes mostly from **successful connections**, not just call attempts.
- **Quality over quantity.**

What Should Telecom Companies Do?

Focus on smart and efficient outreach:

- Use uplift model to identify users with the highest causal responsiveness.
- Limit call attempts to 3–4 per user.
- Aim for longer—but not excessive—calls (up to 200 seconds).
- Prioritize successful contact over mass dialing.

Final Takeaways

- We found outbound calls have a clear causal effect—but only if users are reached (connected). More calls don't always help—focus on smart targeting.
- The findings offer clear suggestions for smarter outreach strategy.
- Our approach blends causal inference + ML for better business insights.

GitHub:

`https://github.com/liangjieliu-dev/unicom-causal-study`

For further technical and empirical details, please refer to the full report.

Thank You!