Causal Inference on the Effectiveness of Outbound Calling Strategies:

Evidence from China Unicom

Submitted by

LIU, Liangjie
Student ID: 21094937

Supervisors

Dr. Min LIN
Ms. Huifang LIU

Master of Science in Data-Driven Modeling

The Hong Kong University of Science and Technology

May 18, 2025

Abstract

This study investigates the causal impact of outbound call strategies on user subscription rates in the telecommunications sector. Leveraging observational data from China Unicom, we employ a hybrid causal inference framework that integrates Propensity Score Matching (PSM) with machine learning-based confounder identification using dual XGBoost models and SHAP values. Our findings show that successful outbound contact—particularly moderate call frequency and mid-range call durations—significantly increases subscription likelihood, though with diminishing returns beyond optimal thresholds. The results highlight the importance of engagement quality over sheer volume and suggest that current outreach strategies may misallocate resources. Practical recommendations are provided to enhance targeting efficiency and maximize marketing ROI.

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

Keywords: causal inference, machine learning, marketing, PSM, SHAP, XGBoost

1 Introduction

In an increasingly saturated telecommunications market, outbound marketing calls remain a common tool to drive user subscriptions. However, their true causal effectiveness is often unclear. Simply increasing call frequency or duration may not yield better results, especially when targeting is suboptimal or user receptiveness is low.

Evaluating such strategies is challenging due to confounding: users more likely to be contacted may also be inherently more likely to subscribe, leading to biased estimates. We develop a causal inference framework that combines Propensity Score Matching (PSM) with machine learning techniques for high-dimensional confounder control.

This study focuses on estimating the causal impact of outbound calls on subscription outcomes, incorporating both direct and mediated effects via call connection. We also examine how call intensity—frequency and duration—influences conversion, offering actionable insights for optimizing outreach.

2 Research Questions and Objectives

This study addresses three central questions: (1) Does outbound call frequency causally increase user subscription rates, and are there diminishing returns? (2) Is call duration a meaningful determinant of conversion? (3) Are observed effects primarily driven by successful connections, or does the act of dialing itself exert an independent influence?

To answer these questions, we (i) build a causal framework to estimate both total and mediated effects of outbound calling; (ii) identify and adjust for high-dimensional confounding using SHAP-informed machine learning; and (iii) estimate average and marginal treatment effects to evaluate call effectiveness under varying intensities.

3 Related Work

Estimating causal effects from observational data poses persistent challenges across disciplines, including economics, healthcare, and marketing. Propensity Score Matching (PSM), introduced by Rosenbaum and Rubin [1], remains a widely used method to reduce selection bias by balancing covariates between treated and control groups.

Traditional PSM struggles with high-dimensional, non-linear relationships [2]. Variable selection is also often manual or heuristic-driven, risking omitted confounders and biased estimates [3].

To overcome these limitations, recent work has integrated machine learning (ML) into causal workflows. Models like XGBoost [4] offer flexibility in capturing complex patterns, while SHAP values [5] provide interpretable measures of variable importance—useful for confounder identification.

However, the combination of SHAP-based feature selection with PSM remains underutilized in applied settings like telecom marketing. Our study addresses this gap by proposing a SHAP-guided confounder pipeline tailored to high-dimensional behavioral data.

Related work by Lo and Li [6] further motivates our approach. They demonstrate how uplift modeling and causal decomposition—explicitly distinguishing between call attempts and successful contact—can yield more targeted and interpretable treatment strategies in outbound marketing.

4 Methodology

To estimate causal effects from observational data, we implemented a multi-step analytical framework. First, we developed a causal model using a Directed Acyclic Graph (DAG) to formalize the assumed relationships between variables. Second, potential confounders were identified using dual XGBoost models and SHAP value analysis. Third, propensity scores were estimated and used for 1:1 nearest neighbor matching with caliper adjustment. Finally, treatment effects were estimated, including average, marginal, and heterogeneous effects, to generate practical marketing recommendations.

4.1 Causal Framework and Assumptions

Our analysis is grounded in the potential outcomes framework and formalized using a Directed Acyclic Graph (DAG) that outlines the assumed causal relationships. As shown in Figure 1, user characteristics and behavioral history—denoted by **C**—are assumed to affect outbound call assignment (T), connection success (A), and eventual subscription decisions (Y).

- T: Whether the user was targeted by outbound calls (intent to treat);
- A: Whether the call attempt resulted in a successful connection (actual exposure);
- Y: Whether the user subscribed to the promoted plan (outcome).

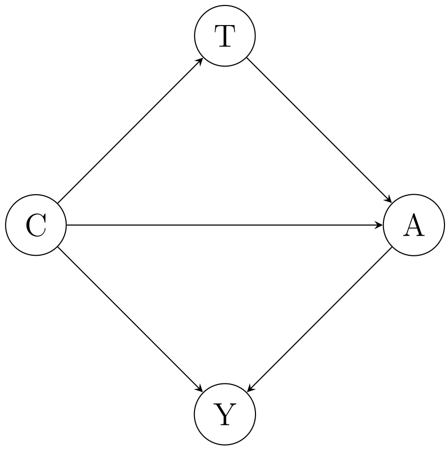

Figure 1: Extended DAG with intermediate call engagement (A).

We rely on standard assumptions from the potential outcomes framework (Appendix A) and apply propensity score methods to adjust for confounding in T, incorporating A as an intermediate variable to separate direct and indirect effects.

4.2 Confounder Identification via Machine Learning and SHAP

To identify confounding variables in high-dimensional user data, we applied a dual-model strategy using gradient-boosted trees (XGBoost) paired with SHAP (SHapley Additive exPlanations). Two separate XGBoost models were trained: one to predict treatment assignment (e.g., likelihood of receiving high-frequency or long-duration calls), and the other to predict subscription outcomes.

XGBoost was selected for its ability to model complex, non-linear interactions and handle missing values. However, its "black box" nature limits interpretability. To address this, we used SHAP values to decompose predictions and quantify each feature's contribution. SHAP offers consistency and local accuracy in attributing feature importance, making it particularly suitable for tree-based models.

We computed SHAP values across both models and selected features that were highly ranked in both treatment and outcome prediction tasks. These intersecting variables were flagged as potential confounders, as they influence both treatment exposure and outcomes. This approach provides an interpretable, data-driven alternative to manual variable selection, improving robustness in high-dimensional settings.

4.3 Propensity Score Modeling and Matching

To adjust for confounding, we used propensity scores estimated from SHAP-selected covariates via logistic regression. Each treated unit was matched to a control unit using 1:1 nearest neighbor matching within a 0.2 caliper on the logit of the propensity score.

Covariate balance was evaluated by comparing standardized mean differences (SMD) before and after matching. All covariates achieved post-matching SMDs below the 0.1 threshold, indicating strong covariate balance. Full formulas are provided in Appendix B.

4.4 Causal Effect Estimation

We adopt the potential outcomes framework to estimate treatment effects, focusing on the **Average Treatment Effect on the Treated (ATT)** as the primary estimand. Since propensity score matching (PSM) constructs a comparison group centered on treated units, ATT provides the most interpretable measure of causal impact in this context.

Uncertainty around ATT estimates was quantified using non-parametric bootstrapping (Appendix C).

In addition to average effects, we also explored marginal and heterogeneous treatment effects by varying treatment thresholds and analyzing subgroup responsiveness. These supplementary analyses provide practical insights into dose-response dynamics and inform more targeted outbound strategies.

4.5 Mediation Analysis Framework

To further understand the mechanism behind the treatment effect, we consider call connection (A) as a mediator between outbound targeting (T) and subscription outcome (Y). This enables us to decompose the total effect into:

- **Direct Effect (DE)**: The effect of being targeted (T) on subscription (Y) that is not explained by connection (A);
- **Indirect Effect (IE)**: The portion of the effect that operates through successful connection, i.e., $T \to A \to Y$.

We estimate the indirect effect using the product-of-coefficients method:

$$\mathrm{IE} \approx [\mathbb{P}(A = 1 \mid T = 1) - \mathbb{P}(A = 1 \mid T = 0)] \times \mathrm{ATT}_{A \to Y} \tag{1}$$

This expression captures how increased contact (via calling) leads to higher subscription likelihood, allowing decomposition into direct and indirect pathways.

This simplified framework follows the logic of counterfactual mediation analysis and provides an interpretable first-order approximation under standard causal assumptions. More advanced methods (e.g., structural equation modeling or g-computation) can extend this analysis in future work.

5 Data Description

This study draws on user-level data from China Unicom's outbound marketing campaigns conducted between December 2024 and January 2025. The dataset encompasses user demographics, behavioral histories, call engagement, and subscription outcomes.

- **Sample Size**: Around 300,000 users, with 870,000 outbound call attempts, including 80,000 successful connections.
- **Timeframe**: December 2024 to January 2025, capturing user behavior and campaign effects over two months.
- **Key Variables**:
  - **User profiles**: Age, gender, customer type, plan type, subscription tenure, recharge behavior, service suspensions, and network usage patterns.
  - **Usage and billing**: Voice/data/SMS consumption, out-of-bundle fees, monthly charges, and indicators of quota overuse or bundling.
  - **Treatment exposure**: Number of outbound call attempts and successful connections, reflecting marketing intensity and reach.
  - **Outcome**: A binary flag indicating whether the user subscribed to the promoted plan, along with plan type and category.

6 Empirical Results

6.1 Unadjusted Bias Analysis

To assess the need for causal adjustment, we begin by examining the full unmatched sample using three dimensions of outbound marketing engagement: **call frequency** (total number of outbound call attempts), **call success count** (number of successfully connected calls), and **call duration** (total time connected in seconds).

Table 1 summarizes the descriptive statistics of these metrics. On average, each user received approximately 2.84 outbound calls, but the median number of successful connections and call duration was zero, implying that over half of the users were never successfully reached. The 80th percentile shows just one successful call and zero seconds of total duration, and only at the 90th percentile does any meaningful hold time emerge. These results show engagement is heavily right-skewed: most users had minimal interaction, while a few dominated contact volume.

Table 1: Descriptive Statistics of Outbound Call Behavior

Statistic	Call Count	Call Success Count	Call Duration (s)
Mean	2.84	0.26	7.73
Std. Dev.	4.83	0.58	40.58
Median	1	0	0.00
70th Pctl	4	0	0.00
80th Pctl	6	1	0.00
90th Pctl	7	1	2.00
Max	58	14	2317

Next, we assess whether any degree of engagement correlates with increased subscription rates. Binary indicators were created for each metric (1 if the value was greater than zero, 0 otherwise). Table 2 compares subscription rates between treated and untreated users across these engagement dimensions.

Table 2: Subscription Rates by Call Activity (Any vs. None)

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

The differences are stark. Users who received at least one outbound call were over three times more likely to subscribe than those who received none (8.65% vs. 2.23%). The effect was even stronger when considering actual connection: users who answered at least one call exhibited a 17.40% conversion rate compared to just 2.39% among those never reached. Similarly, users with any measurable call duration showed a subscription rate of 11.86%, more than double that of users with zero engagement.

These unadjusted comparisons clearly suggest a strong association between deeper engagement and conversion. However, they remain correlational in nature: users who are easier to reach or more actively targeted may differ systematically in characteristics that also affect their likelihood to subscribe. Thus, these results must be interpreted cautiously, and causal inference techniques—such as propensity score matching—are required to disentangle correlation from true treatment effects.

6.2 Identified Confounders

To identify confounding variables that influence both treatment assignment and outcomes, we applied a dual-model SHAP approach. Specifically, we trained two separate XGBoost classifiers:

- A **treatment model** to predict whether a user received any successful outbound call (`call_success_cnt > 0`);
- An **outcome model** to predict whether the user subscribed to the promoted product (`is_order = 1`).

SHAP values were used to measure each feature's contribution to the model. We ranked the top 15 features from each model based on their mean absolute SHAP values and identified the intersection set—12 features that were highly predictive of both treatment and outcome. These intersecting variables—spanning demographics, usage, and plan features—were incorporated into the PSM model.

Full feature rankings and the selected confounder set are documented in Appendix D (Tables 5–6, Figures 3–4).

6.3 Matching Quality Assessment

To address confounding, we conducted propensity score matching (PSM) using the 12 confounders identified via SHAP analysis. The objective was to balance observed covariates between users who had successful outbound call engagement and those who did not.

Covariate balance was evaluated using **Standardized Mean Differences** (SMDs), with 0.1 as the threshold for acceptable balance. Figure 5 shows that substantial imbalance existed prior to matching, particularly for `cust_type` (SMD = 0.399) and `m_cur_income` (SMD = 0.274). After matching, all covariates achieved SMDs below 0.1, with most falling under 0.05.

This indicates that the matching procedure was effective in constructing comparable treatment and control groups, thereby supporting the validity of causal inference.

A detailed breakdown of covariate balance before and after matching is presented in Appendix E (Table 7), with a visual summary shown in Figure 5.

6.4 Causal Effect Estimates

To assess the causal impact of outbound call activity on subscription behavior, we estimated the Average Treatment Effect on the Treated (ATT) using propensity score matching (PSM) under three binary treatment definitions: receiving any call, having at least one successful connection, and accumulating any call duration.

Binary Treatment Effects. As shown in Table 3, all forms of contact exert statistically significant causal effects. The most substantial uplift comes from **successful call connections**, with an ATT of 15.24%. In contrast, merely placing outbound calls without ensuring connection yields a smaller effect (6.64%), while any amount of call duration results in an ATT of 8.97%.

These findings indicate that **conversion is driven primarily by actual engagement**, not just contact attempts. This insight emphasizes the need to prioritize contact quality over volume.

Table 3: Estimated Causal Effects from Propensity Score Matching (95% Confidence Intervals)

Treatment Type	Unmatched Diff	ATT (Matched)	95% Confidence Interval	Bias Reduction
Call Frequency (`call_cnt > 0`)	6.42%	6.64%	(6.49%, 6.79%)	−0.22 pp
Call Success (`call_success_cnt > 0`)	15.01%	15.24%	(14.91%, 15.56%)	−0.22 pp
Call Duration (`call_hold_time > 0`)	7.09%	8.97%	(8.58%, 9.38%)	−1.89 pp

Marginal (Dose-Response) Effects. To examine how treatment intensity affects outcomes, we varied threshold levels and observed marginal changes in conversion. Results show clear **diminishing returns** across all dimensions. For instance, the first successful call delivers the largest benefit, while additional calls yield less incremental value. Likewise, optimal call duration appears to fall between 50–200 seconds.

The full dose-response curves are presented in Appendix F (Figures 6–8).

Mediation Analysis: Does Connection Drive the Effect of Calling? To quantify how much of the overall treatment effect is explained by successful contact, we apply a simple mediation decomposition based on observed probabilities. Table 4 presents the relevant quantities.

Table 4: Mediation Effect Decomposition of Outbound Calling

Quantity	**Value**
Connection Rate ($P(A = 1 \mid T = 1)$)	0.4075
Conversion if Connected ($P(Y = 1 \mid A = 1, T = 1)$)	0.1740
Conversion if Not Connected ($P(Y = 1 \mid A = 0, T = 1)$)	0.0264
Pickup Effect (Δ_{pickup})	0.1476
Indirect Effect	**0.0602**
Total Effect ($ATT_{T \to Y}$)	**0.0664**
Direct Effect	**0.0062**

Figure 2 visualizes the decomposition. The results show that over 90% of the total effect of outbound calling on subscription is mediated by successful contact. This highlights that merely dialing users has minimal impact unless the call is actually answered.

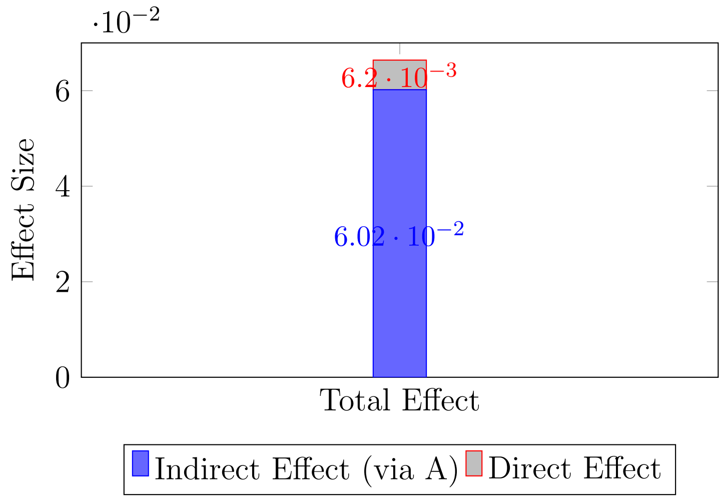

Figure 2: Decomposition of Total Effect of Calling on Subscription

6.5 Empirical Findings and Strategic Implications

1. Resource Misallocation in Current Outreach Strategy Across all treatment definitions, the **matched ATT consistently exceeds the unadjusted differences** in subscription rates, accompanied by negative bias values. This indicates that the current outbound marketing strategy **does not effectively prioritize users who are most responsive to contact**, leading to potential inefficiencies in resource allocation.

More specifically, the current assignment mechanism appears to rely on static characteristics—such as historical subscription propensity—rather than **causal responsiveness to outreach**. Consequently, marketing resources may be disproportionately directed toward low-uplift users, while genuinely persuadable users are overlooked.

To address this shortcoming, we recommend implementing an **uplift modeling approach** that estimates individual-level treatment effects. By identifying users with the highest causal responsiveness—based on behavioral signals, usage profiles, and commercial value—telecom operators can adopt a **uplift-based targeting strategy** that maximizes ROI and minimizes inefficiency. Recent work by Lo and Li [6] provides an advanced framework for this objective.

2. Insights from Marginal Effects The dose-response plots (Figures 6–8) reveal several key operational patterns:

- **Successful Calls:** The first successful connection yields the largest marginal gain. Subsequent calls show rapidly diminishing returns, suggesting that *a single effective conversation is often sufficient.*
- **Call Attempts:** Subscription lift increases with up to three attempts but declines beyond seven, highlighting that *excessive contact risks user fatigue without additional benefit.*

- **Call Duration:** Conversations lasting 50–200 seconds are associated with the highest conversion probabilities. Beyond this range, the effect plateaus and becomes unstable, indicating that *call quality outweighs call length.*

These findings emphasize the **nonlinear and diminishing returns of outbound engagement**. To enhance both efficiency and user experience, outbound strategies must carefully calibrate contact frequency, timing, and depth of interaction.

3. Insights from Mediation Analysis The mediation analysis reveals that the bulk of the causal impact of outbound calls on subscription is not due to the act of calling itself, but rather whether the user successfully connects. Specifically, 90.7% of the total treatment effect (0.0602/0.0664) can be attributed to the indirect pathway via call connection, while only 9.3% is explained by the direct effect of call attempts that did not result in contact.

This finding reinforces a key operational insight: **connection quality, not contact volume, drives conversions**. Attempting to reach users without ensuring call success has limited causal impact. As such, dialing strategies that maximize actual connection—such as smarter timing, repeated but spaced retries, or alternate channels—are likely to yield significantly better returns than simply increasing outbound volume.

7 Limitations and Future Work

- **Regional Scope:** Data is from a single region; future work should test generalizability.
- **Unobserved Confounding:** PSM controls observed factors, but hidden biases may remain; methods like IV or DML could help.
- **PSM Constraints:** IPW or doubly robust methods are worth exploring.
- **Narrow Treatment Context:** Focus is on basic subscriptions; future studies could explore upgrades, retention, or adaptive strategies.

8 Conclusion and Implications

Moderate call frequency and mid-length duration significantly improve subscription rates. Engagement quality matters more than volume.

The proposed framework—combining PSM with SHAP-informed feature selection—offers a practical and interpretable approach for causal inference in marketing.

Future work can validate findings in broader contexts or integrate causal models into adaptive outreach systems.

References

[1] Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[2] Guido W Imbens and Donald B Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press, 2015.

[3] Susan Athey and Guido W Imbens. The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2):3–32, 2017.

[4] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM, 2016.

[5] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, volume 30, pages 4765–4774, 2017.

[6] Victor S. Y. Lo and Zhuang Li. Causal inference and machine learning techniques for outbound calls strategy. In *Proceedings of the Joint Statistical Meetings (JSM)*. American Statistical Association, 2021. Presented by Fidelity Investments.

Appendix A: Assumptions for Causal Identification

Causal inference using observational data requires several standard assumptions from the potential outcomes framework. These are not directly testable but provide the foundation for interpreting estimated effects as causal:

- **Stable Unit Treatment Value Assumption (SUTVA)**: The treatment or outcome of one individual is unaffected by the treatment assignment of others, and there is only one version of each treatment.

- **Unconfoundedness (Conditional Independence)**: Given observed covariates $\mathbf{C}$, treatment assignment T is independent of potential outcomes. That is, $Y(t) \perp\!\!\!\perp T \mid \mathbf{C}$ for $t = 0, 1$.

- **Overlap (Common Support)**: Every unit has a non-zero probability of receiving each treatment condition, i.e., $0 < \mathbb{P}(T = 1 \mid \mathbf{C}) < 1$.

These assumptions underpin the use of propensity score methods and mediation decomposition in this study.

Appendix B: Technical Definitions for Matching

Propensity Score. The propensity score is defined as the conditional probability of receiving treatment given covariates:

$$e(X) = \mathbb{P}(T = 1 \mid X)$$

Standardized Mean Difference (SMD). To assess covariate balance before and after matching, we use:

$$\text{SMD} = \frac{\bar{X}_T - \bar{X}_C}{\sqrt{\frac{1}{2}(s_T^2 + s_C^2)}}$$

An SMD less than 0.1 indicates adequate balance.

Appendix C: Bootstrap Procedure for ATT Estimation

1. Given n matched treated-control pairs, repeat the following process $B = 1000$ times:
   - Sample n pairs with replacement from the matched dataset.
   - For each bootstrap sample, compute the ATT as:

$$\widehat{\text{ATT}}^{(b)} = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i^{(\text{Treated})} - Y_i^{(\text{Control})} \right)$$

2. Construct a 95% confidence interval by extracting the 2.5th and 97.5th percentiles from the empirical distribution of $\{\widehat{\text{ATT}}^{(1)}, \ldots, \widehat{\text{ATT}}^{(B)}\}$.

Appendix D: SHAP Feature Rankings and Selected Confounders

A.1 Selected Confounders (Top-Ranked Intersection)

Table 5 lists the 12 variables that were identified as high-priority confounders, appearing in the top 15 features of both the treatment and outcome models.

Table 5: Intersection Set of SHAP-Selected Confounders

Variable Name	**Description (Short)**
`age`	User age
`cust_type`	Customer type
`user_state_codeset`	Account status
`join_month`	Subscription tenure
`m_add_duration_01`	Monthly call duration
`m_add_call_times`	Outbound call times
`m_num_bill_duration`	Billed call duration
`product_id`	Product type
`product_rent`	Product rental fee
`m_add_gs_234g_net`	Network capability (2/3/4/5G)
`m_sw_status_month`	5G switch indicator
`m_cur_income`	Current revenue level

A.2 Top 15 Features from Treatment and Outcome Models

Table 6: Top 15 SHAP Features from Treatment and Outcome Models

Treatment Feature	**SHAP Value**	**Outcome Feature**	**SHAP Value**
m_add_call_times	0.239	m_cur_income	0.356
cust_type	0.205	m_total_saturation	0.214
product_rent	0.189	user_state_codeset	0.171
m_is_umlimited	0.157	product_id	0.161
m_add_duration_01	0.153	cust_type	0.137
m_cur_income	0.149	m_add_duration_01	0.137
user_state_codeset	0.122	join_month	0.106
m_add_call_times_01	0.117	m_sw_status_month	0.097
m_sw_status_month	0.112	age	0.096
m_add_gs_234g_net	0.099	product_rent	0.095
product_id	0.082	m_add_call_times	0.093
age	0.082	m_num_bill_duration	0.090
join_month	0.073	m_add_gs_234g_net	0.083
m_mealout_call_fee	0.046	m_add_duration	0.082
m_num_bill_duration	0.029	develop_depart_id	0.077

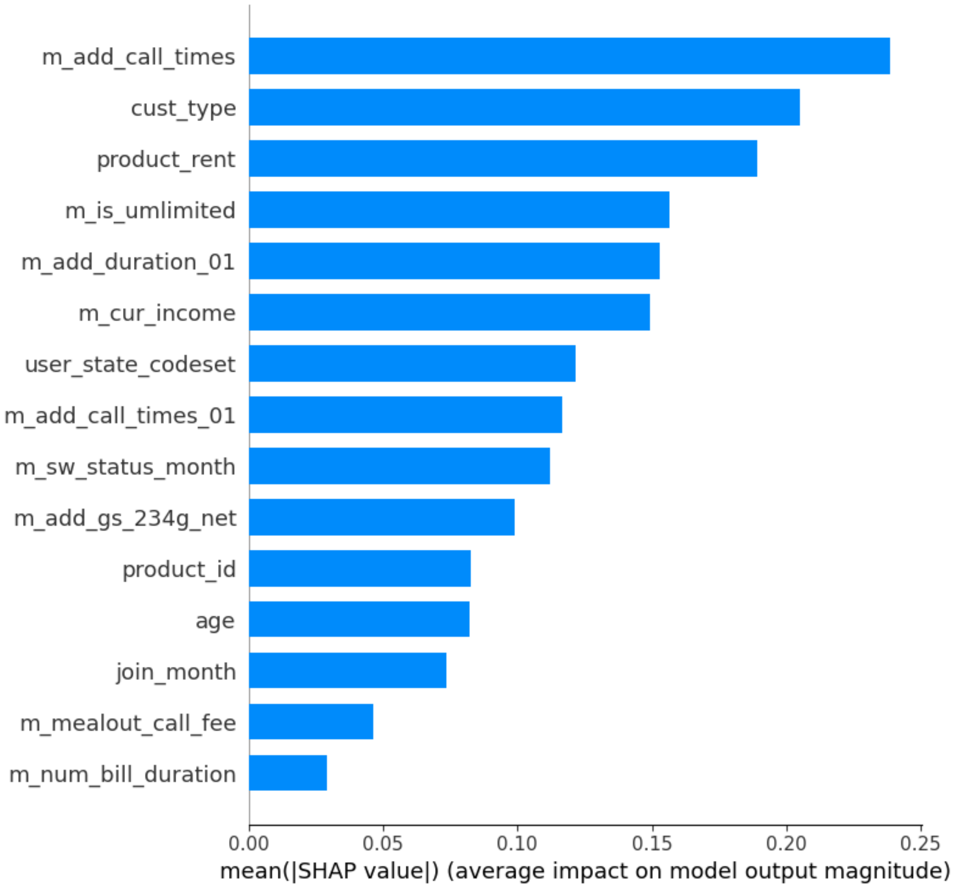

Figure 3: SHAP Summary Plot – Treatment Model

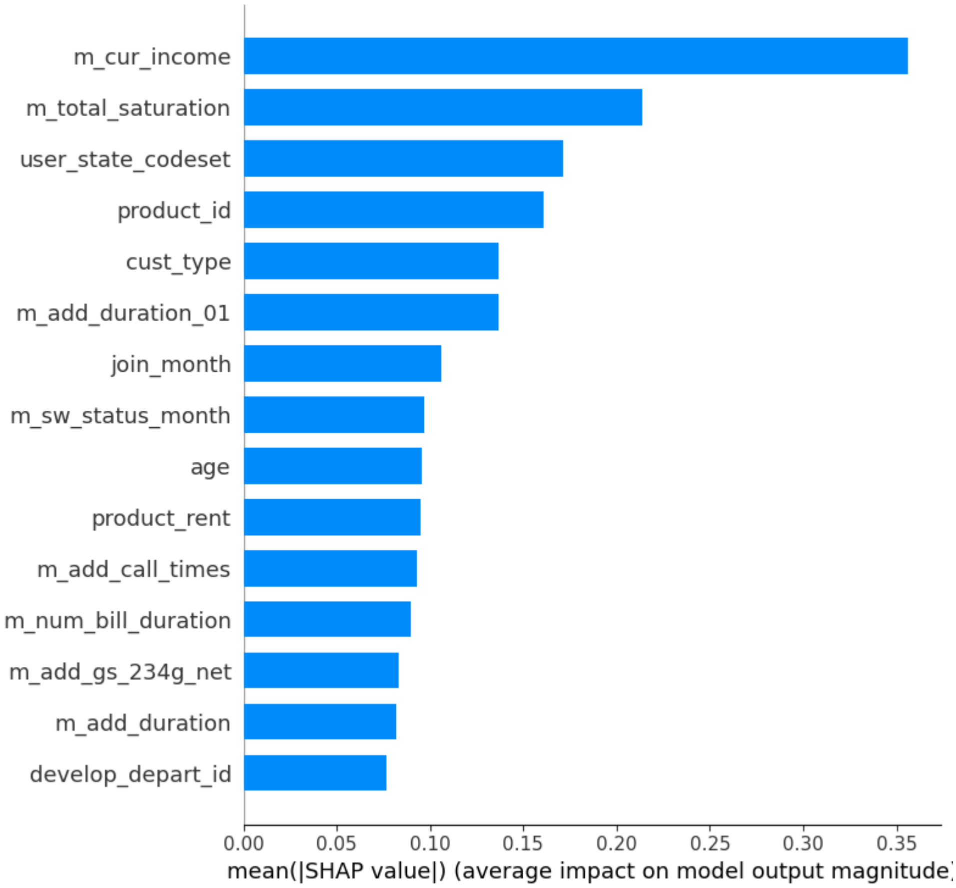

Figure 4: SHAP Summary Plot – Outcome Model

Appendix E: Covariate Balance and SMD Analysis

Table 7: Standardized Mean Differences Before and After Matching

Covariate	SMD (Before)	SMD (After)
cust_type	0.399	0.0009
m_sw_status_month	0.346	0.0524
m_cur_income	0.274	0.0594
user_state_codeset	0.260	0.0092
product_rent	0.215	0.0406
product_id	0.192	0.0089
m_add_gs_234g_net	0.174	0.0279
age	0.114	0.0097
join_month	0.095	0.0189
m_num_bill_duration	0.078	0.0326
m_add_call_times	0.057	0.0295
m_add_duration_01	0.052	0.0183

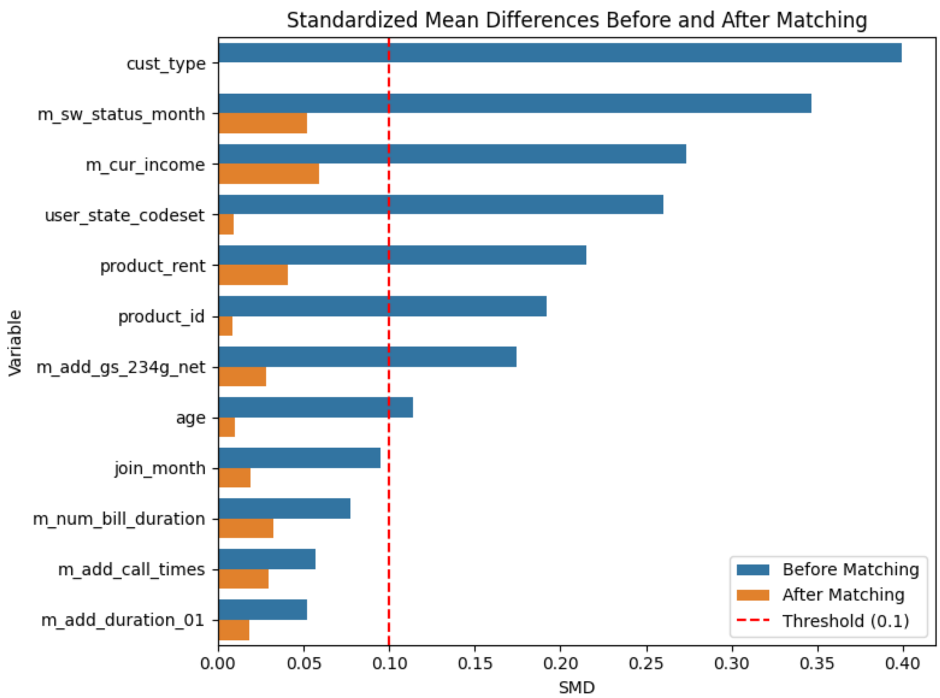

Figure 5: SMD Before and After Matching

Appendix F: Marginal (Dose-Response) Effects

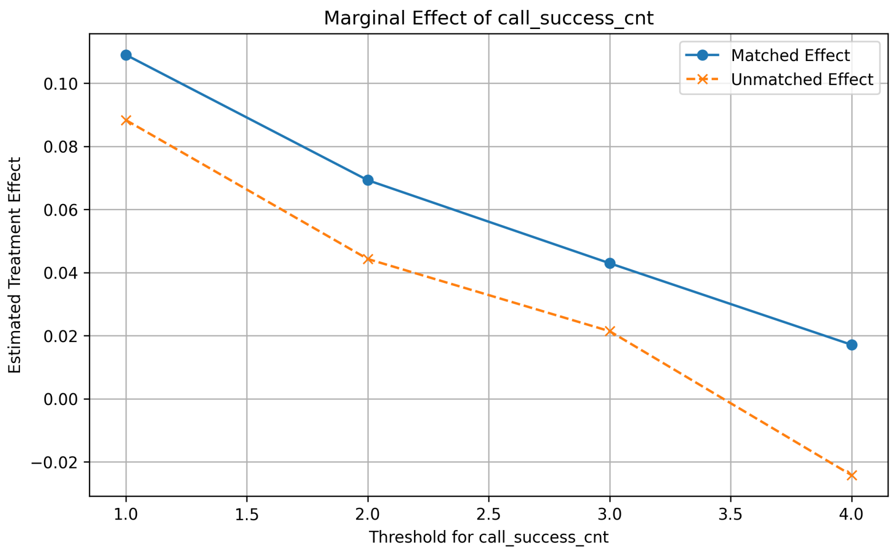

Figure 6: Marginal Treatment Effect by Call Success Count Threshold

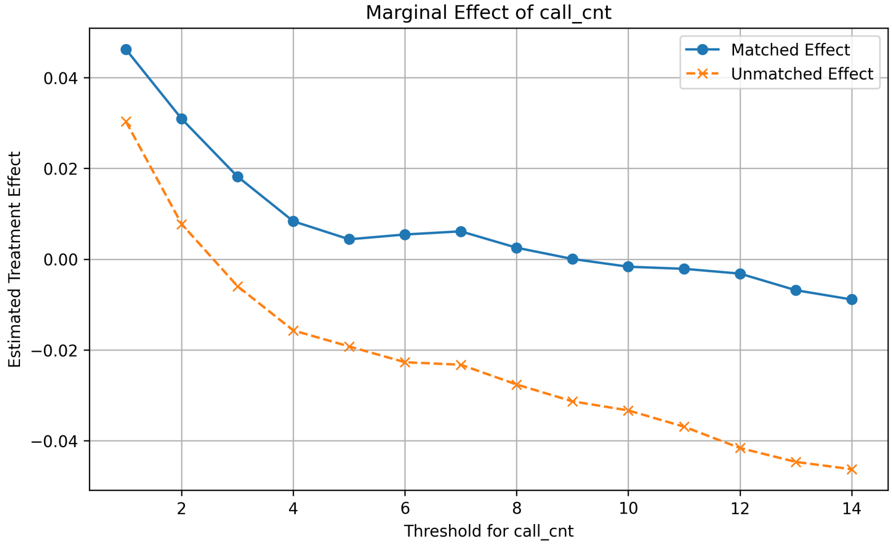

Figure 7: Marginal Treatment Effect by Call Frequency Threshold

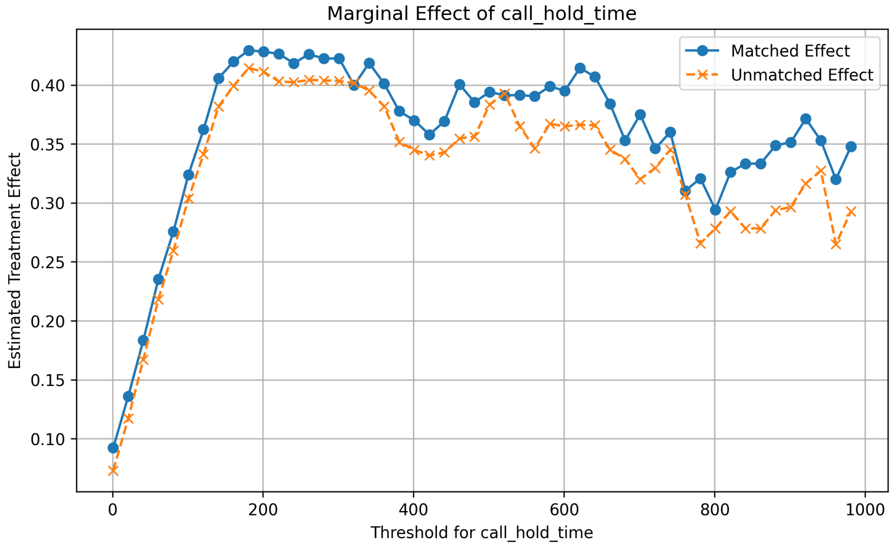

Figure 8: Marginal Treatment Effect by Call Duration Threshold