

# **An Empirical Analysis of Calendar Effects and Macroeconomic Trends on Marketing Campaign Outcomes**

A Project for the Course “Machine Learning for Econometrics”

**Elea  
Bordais**  
(ENSAE)

**Jade  
Sillere**  
(IP-Paris)

**Eliott  
Von-Pine**  
(ENSAE)

**Patryk  
Wiśniewski**  
(ENSAE)

## **Abstract**

This report presents actionable insights to enhance cost-effectiveness in marketing campaigns by analyzing data from a bank marketing initiative. It examines the impact of timing on consumer engagement, providing credible estimates of differential responses to the products offered. Additionally, the report offers evidence of the significance of macroeconomic context and broad market trends in the success of campaigns. It further investigates consumer heterogeneity and its effects on marketing effectiveness and time preference. Finally, the report serves as a reference for comparing traditional and modern machine learning methods for causal identification.

# 1 Introduction and Research Question

Telemarketing, as defined by the Mobile Marketing Association, involves practices that enable organizations to engage interactively and relevantly with audiences via mobile devices or networks. In the banking sector, telemarketing has been widely utilized since the 1980s due to its cost-efficiency and capacity for personalized customer engagement. Banks established call centers and developed consumer databases to transition away from reliance on in-person sales, enabling targeted promotion of products such as loans, credit cards, and term deposits.

The adoption of telemarketing is driven by several factors, including higher conversion rates compared to mass marketing, rapid customer outreach capabilities, and ease of campaign performance monitoring. However, challenges persist, such as low response rates, customer fatigue, and negative perceptions associated with unsolicited calls. A significant challenge for institutions remains accurately identifying customers likely to subscribe to specific products.

The complexity of telemarketing campaigns makes identifying key success drivers difficult. Rasool Basha 2024 emphasizes factors such as telemarketer skills, call list quality, and call timing. Data-driven approaches offer substantial improvements to telemarketing strategies. For instance, Moro, Cortez, and Rita 2014 analyzed a Portuguese banking dataset to predict telemarketing success rates for term deposit sales. Their study indicated neural networks provided superior predictive accuracy and identified critical predictive factors, including the Euribor rate, call direction (inbound or outbound), and bank agent experience.

Building upon the aforementioned contribution, this report adopts a causal analysis framework to evaluate the impact of timing on telemarketing success rates within the banking industry. Answering the question of what is the best time to reach out to a potential client enables for a more informed decisions of resource allocation.

Recent advancements in causal machine learning, particularly the Double Machine Learning (DML) methodology introduced by Chernozhukov et al. 2018, facilitate addressing causal questions using machine learning tools already familiar within predictive analytics. Consequently, this study seeks to compare and evaluate various methods for identifying causal effects in the context of telemarketing campaign optimization.

## 2 Data

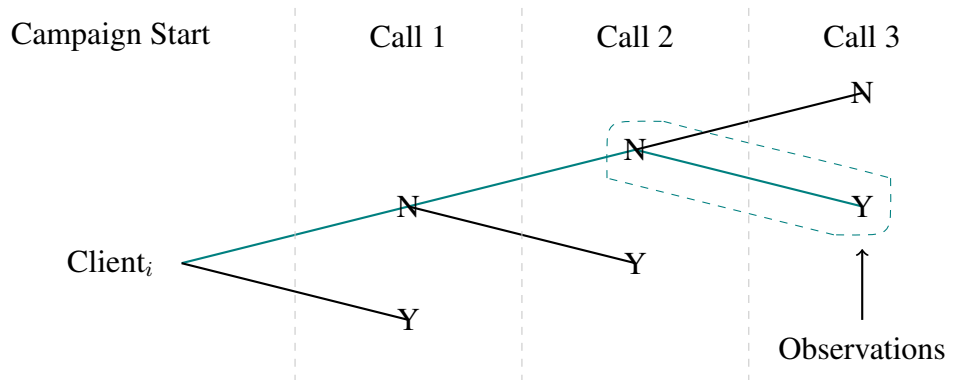
### 2.1 Marketing Campaign

Our primary dataset originates from Moro, Cortez, and Rita 2014, who analyzed data obtained from a bank marketing campaign conducted by a Portuguese bank between 2008 and 2010. The campaign focused on selling term deposits to clients, during a period when the bank did not utilize predictive marketing techniques. The dataset allowed the authors to assess which individuals should be prioritized in marketing efforts.

The dataset comprises approximately 40,000 observations and includes over 20 variables, containing more than 80 features. It encompasses demographic details such as educational levels and employment statuses, financial indicators like loan status, account balances, or defaults, as well as methodological aspects such as contact methods and historical engagement with previous campaigns. Additionally, we complemented the dataset with Portuguese macroeconomic indicators sourced from publicly available data from institutions like the World Bank or Eurostat.

It is important to note that our version of the dataset differs from the original used by Moro, Cortez, and Rita 2014. Our dataset is smaller, censored, and does not contain the full range of individual-level details due to privacy concerns. Additionally, the data is not in a panel structure.

Figure 1: Campaign Process and Observed Data



This dataset records only the final successful interaction with each individual. For instance, if an individual was contacted multiple times, we observe only the outcome of their last interaction. We also do not have direct access to the bank's client dataset. This presents an identification challenge, as individuals who were recontacted after an initial rejection differ, possibly due to unobservable factors (e.g., requested new contact), from those who were contacted only once.

## 2.2 Descriptive Statistics

We present descriptive statistics to provide insights into the characteristics of our dataset.

Table 1: Observed Outcomes

Outcome	Count	Share (%)
Failure	36548	88.73
Sucess	4640	11.27

Firstly, we observe that the outcome variable is significantly imbalanced, exhibiting a failure rate of approximately 89%. Additionally, as previously noted, individuals who receive multiple contacts demonstrate a decreasing probability of acceptance with each subsequent call.

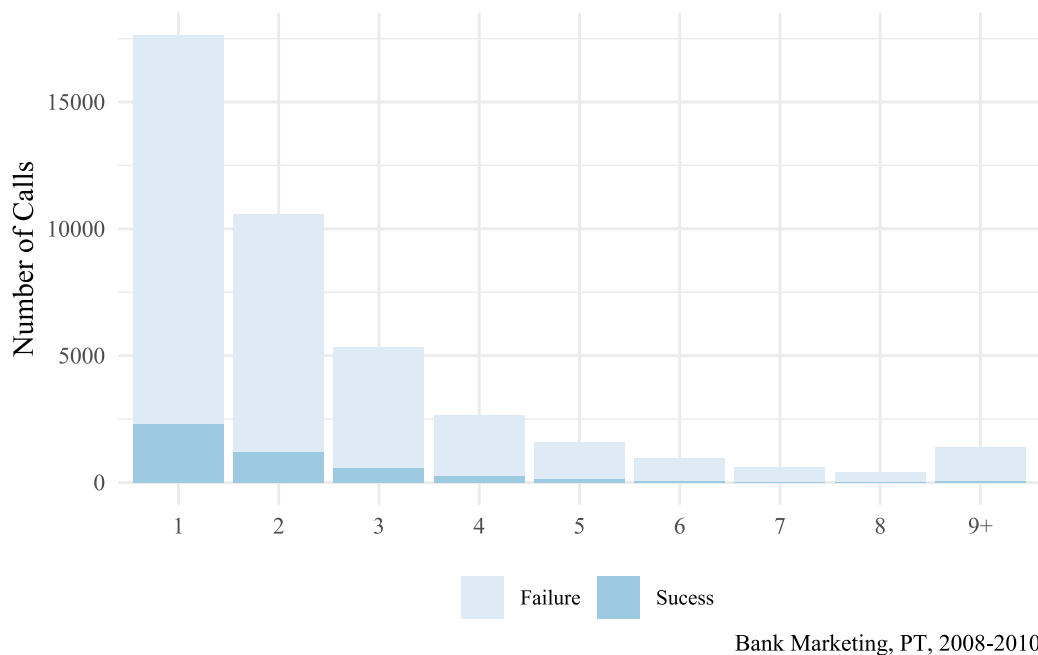


Figure 2: Probability of sucess and number of calls

However, it is important to highlight that nearly half of the observations consist of individuals contacted only once. This initial contact should remain unaffected by unobserved confounders. Consequently, our analysis will primarily focus on these first-contact interactions.

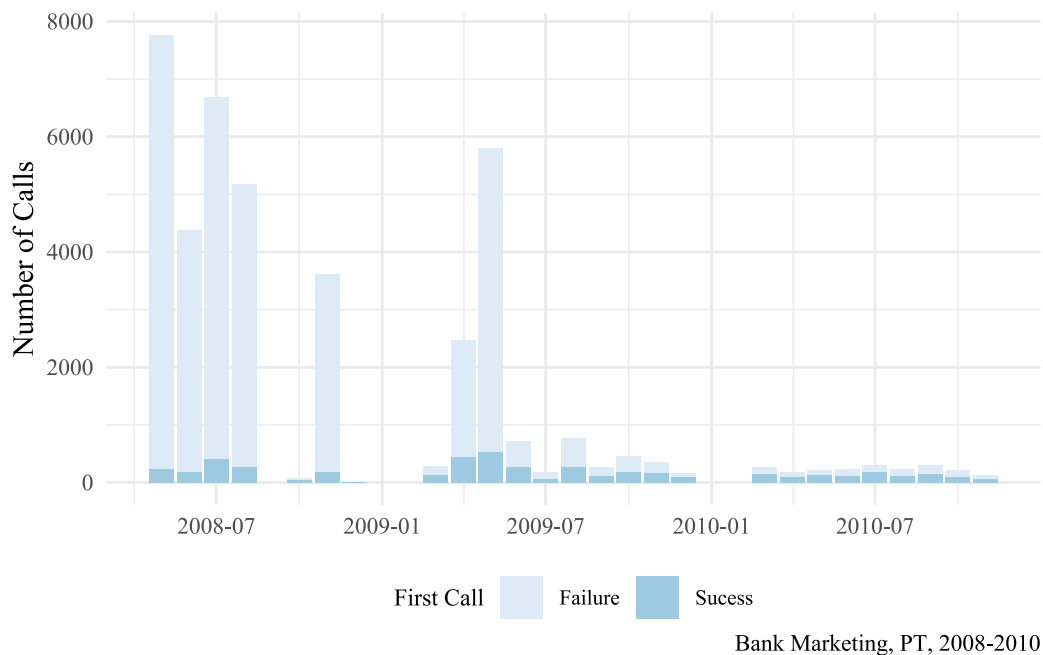


Figure 3: Number of observed last calls

Examining the temporal distribution of the campaign, we find that marketing efforts were not evenly distributed over time. The majority of calls occurred early in the campaign, tapering off substantially towards 2010. This trend may reflect strategic shifts by the bank following the onset of the Great Financial Crisis, potentially driven by changes in interest rates affecting product profitability or consumer demand.

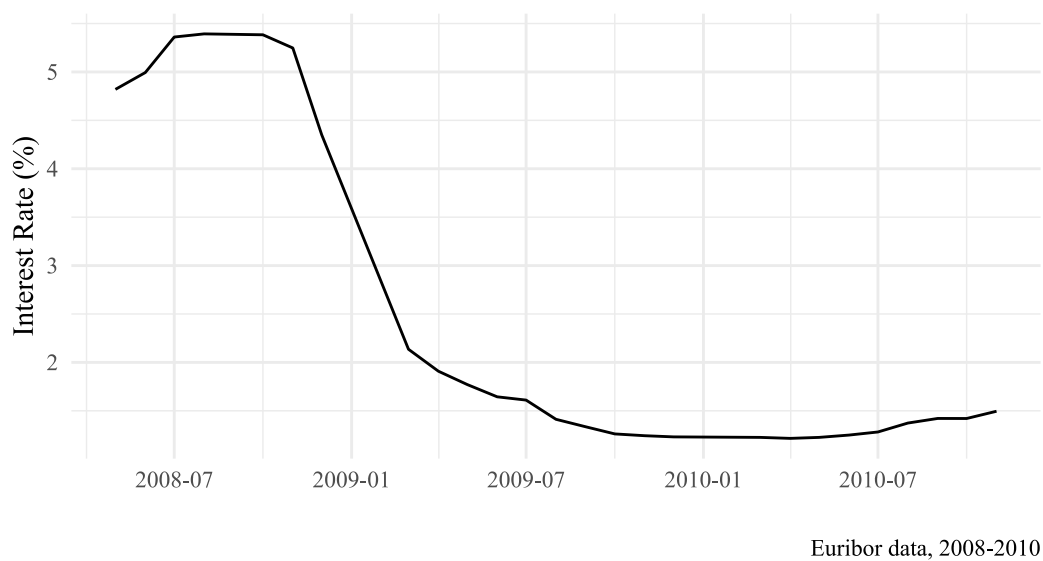


Figure 4: Euribor 12mo rates

Nonetheless, the variation in interest rates over this period provides an opportunity to investigate their impact on take-up rates, which will be explored in details in the final section. A preliminary correlation analysis reveals, as anticipated, a negative correlation between the number of contacts and acceptance probability during the last contact.

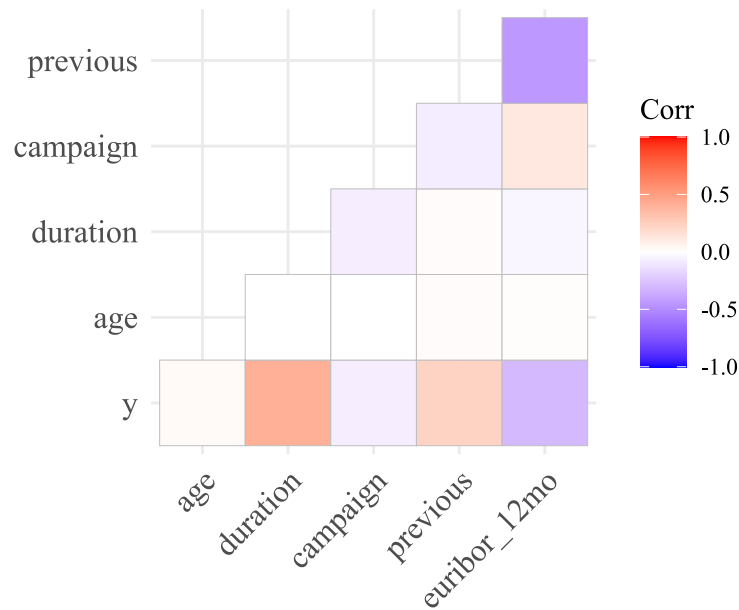


Figure 5: Correlations between numeric variables

Furthermore, we observe a negative correlation between interest rates and acceptance rates. This relationship will be analyzed more comprehensively in the subsequent section but is likely indicative of demand shocks.

Table 2: Share of successful calls

Share	Monday	Tuesday	Wednesday	Thursday	Friday
Failure	0.901	0.882	0.883	0.879	0.892
Success	0.099	0.118	0.117	0.121	0.108
Count	8514.000	8090.000	8134.000	8623.000	7827.000

Finally, the data indicates that the day of the week may influence the likelihood of a successful outcome.

## 2.3 Imputations (MICE)

The dataset contains several “unknown” entries, however, given the significant reduction in sample size resulting from selecting only single-contact observations, further exclusion of data points is undesirable. Therefore, we opted to impute missing values.

Table 3: Number of ‘unknown’ observations

job	marital	education	default	housing	loan
158	36	740	3481	415	415

Missing data can introduce bias and compromise the accuracy of results. We assume that the missing values follow a Missing at Random (MAR) pattern, meaning the probability of a value being missing depends solely on observed data rather than on unobserved variables. This means that after accounting for available data, “any remaining missingness is completely random.” (Graham 2009). This permits the use of Multiple Imputation by Chained Equations (MICE), a method that models relationships between variables to generate plausible values that mirror observed patterns of missingness. The procedure can be summarized in three key phases:

1. **Imputation Phase:** Each variable containing missing values is imputed based on other variables. MICE employs a regression-based approach, predicting missing values from observed data for each incomplete variable sequentially.
2. **Iteration Phase:** The imputation step is iteratively repeated, typically around ten cycles, to allow for convergence of imputed values (Raghunathan 2004).
3. **Pooling Phase:** The imputation procedure is performed multiple times ( $m$  iterations), resulting in several complete datasets. Following established guidelines, we chose to perform five imputation iterations ( $m = 6$ ). This decision aligns with recommendations from Bennett 2001, suggesting that 5 to 10 iterations are appropriate. Hawthorne, Hawthorne, and Elliott 2005 similarly advocate for fewer than ten iterations, while Sterne et al. 2009 specifically recommend three or five iterations.

Regressions from each imputed dataset are combined using Rubin’s rules (Rubin 1987).

### 3 Methodology, Econometrics and Machine Learning

To answer the question of best timing, we employ both traditional econometric methods, such as logistic regression, and advanced nonparametric machine learning techniques. Nonparametric ML approaches are particularly valuable as they effectively capture complex, highly nonlinear relationships. Specifically, we use Double/Debiased Machine Learning (DDML) and Causal Forests.

#### 3.1 Double/Debiased Machine Learning

Double/Debiased Machine Learning (DDML), introduced by Chernozhukov et al. 2018, estimates causal effects in high-dimensional settings where relationships among variables may be nonlinear and complex. DDML integrates econometric inference and ML's flexible predictive modeling, providing reliable causal inference while managing regularization bias typically associated with ML methods.

DDML relies on two critical principles:

We use DDML to estimate the causal effect of a treatment or policy variable on an outcome of interest, while controlling for high-dimensional confounders. In such settings, classical parametric methods may fail if the true relationships are nonlinear or if the relevant control variables are unknown or numerous. Machine learning methods can flexibly model these complex relationships, but they often introduce regularization bias and do not provide valid inference for causal parameters.

DDML relies on two critical principles:

1. Neyman Orthogonality: This ensures that the estimation is robust to small perturbations in nuisance parameter estimates, thus reducing bias.
2. Cross-fitting: Data is partitioned to separately estimate nuisance parameters and causal effects, avoiding overfitting.

This method relies on two key principles : Neyman orthogonality that removes the influence of control variables on both the treatment and the outcome, reducing bias ; and cross-fitting that avoids over-fitting by using different parts of the data to estimate the control functions and the final effect.



Our analysis focuses on two primary models within the DDML framework: the Partially Linear Model (PLM) and the Interactive Regression Model (IRM).

### 3.2 Partially Linear Model (PLM)

The PLM specifies the following relationship:

$$Y = \theta_0 D + g_0(X) + U$$

Here,  $\theta_0$  represents the causal effect of the treatment on the outcome, and captures an unknown, potentially nonlinear relationship involving covariates. A critical assumption for identification is conditional orthogonality:

$$\mathbb{E}[\text{Cov}(U, D \mid X)] = 0$$

Letting  $\ell_0(X) = \mathbb{E}[Y \mid X]$  and  $m_0(X) = \mathbb{E}[D \mid X]$ , the parameter  $\theta_0$  can be identified by the following moment condition:

$$\theta_0 = \frac{\mathbb{E}[(Y - \ell_0(X))(D - m_0(X))]}{\mathbb{E}[(D - m_0(X))^2]}$$

The estimation procedure for the DDML involves two steps:

1. Estimating nuisance functions  $\ell_0(X)$  and  $m_0(X)$  with flexible ML methods (e.g., Lasso, Random Forest, Boosting).
2. Plug in the predicted values to residualize  $Y$  and  $D$ , and compute  $\hat{\theta}_n$  as the sample analogue of Equation using cross-fitting to reduce overfitting bias.

$$\hat{\theta}_n = \frac{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{\ell}^{-i}(X_i))(D_i - \hat{m}^{-i}(X_i))}{\frac{1}{n} \sum_{i=1}^n (D_i - \hat{m}^{-i}(X_i))^2}$$

Here, cross-fitting entails dividing the dataset into  $K$  folds, estimating nuisance functions on one subset and evaluating them on another, significantly reducing the risk of overfitting.

### 3.3 Interactive Regression Model (IRM)

The IRM provides a fully flexible modeling of the treatment-covariate interaction and is specified as:

$$Y = g_0(D, X) + U$$

In this model,  $D$  is a binary treatment indicator. Unlike PLM, IRM does not require additive separability between  $D$  and the covariates  $X$ . We focus on two causal parameters:

- Average Treatment Effect (ATE):

$$\theta_0^{ATE} = \mathbb{E}[g_0(1, X) - g_0(0, X)]$$

- Average Treatment Effect on the Treated (ATET):

$$\theta_0^{ATET} = \mathbb{E}[g_0(1, X) - g_0(0, X) \mid D = 1]$$

IRM identification requires two assumptions:

- **Conditional Mean Independence:**  $\mathbb{E}[U \mid D, X] = 0$
- **Overlap:**  $\mathbb{P}(D = 1 \mid X) \in (0, 1)$  with probability 1.

Let  $g_0(d, X) = \mathbb{E}[Y \mid D = d, X]$  and  $m_0(X) = \mathbb{E}[D \mid X]$  be the nuisance functions. Then, the efficient influence function for the ATE is:

$$\psi^{ATE}(W) = \frac{D(Y - g(1, X))}{m(X)} - \frac{(1 - D)(Y - g(0, X))}{1 - m(X)} + g(1, X) - g(0, X) - \theta$$

The DDML estimator solves the empirical moment condition:

$$\frac{1}{n} \sum_{i=1}^n \psi^{ATE}(W_i; \hat{g}, \hat{m}) = 0 \quad \Rightarrow \quad \hat{\theta}_n^{ATE}$$

Similar to the PLM, nuisance parameters are estimated through ML methods combined with cross-fitting to achieve robustness against regularization bias and overfitting.

### 3.4 Causal Forests

The Causal Forest is a nonparametric machine learning method designed to estimate heterogeneous treatment effects at the individual or subgroup level. Proposed by Wager and Athey 2018, it builds on the classic random forest algorithm but adapts it for causal questions rather than pure prediction. The method relies on a generalized random forest framework, where each tree is grown using a splitting criterion that maximizes differences in treatment effects between subgroups. Causal forests are particularly useful when treatment effects vary across individuals and when the structure of this heterogeneity is unknown. Importantly, they allow for valid inference on individual-level treatment effects under assumptions similar to those of unconfoundedness and overlap.

The method operates within the *potential outcomes framework*, assuming the standard *unconfoundedness* condition:

$$\{Y^{(0)}, Y^{(1)}\} \perp\!\!\!\perp W \mid X$$

and targets the *Conditional Average Treatment Effect (CATE)*:

$$\tau(x) = \mathbb{E}[Y^{(1)} - Y^{(0)} \mid X = x]$$

Each causal tree partitions the feature space and estimates the treatment effect in each leaf as the difference in average outcomes between treated and control observations:

$$\hat{\tau}(x) = \frac{1}{|S_1(x)|} \sum_{i \in S_1(x)} Y_i - \frac{1}{|S_0(x)|} \sum_{i \in S_0(x)} Y_i$$

where  $S_1(x)$  and  $S_0(x)$  are treated and control units in the same leaf as  $x$ .

The forest combines many such trees, each built on subsamples of the data, yielding the final estimator:

$$\hat{\tau}(x) = \frac{1}{B} \sum_{b=1}^B \hat{\tau}_b(x)$$

A key feature of causal forests is the use of *honest trees*, where sample-splitting ensures that the data used to build the tree structure is not reused to estimate treatment effects within leaves. This guarantees asymptotic properties such as consistency and normality. Under regularity conditions, the estimator satisfies:

$$\frac{\hat{\tau}(x) - \tau(x)}{\sqrt{\widehat{\text{Var}}[\hat{\tau}(x)]}} \xrightarrow{d} \mathcal{N}(0, 1)$$

The variance can be consistently estimated using the *infinitesimal jackknife* method. Overall, causal forests provide a flexible and statistically rigorous tool for personalized causal inference, especially in high-dimensional or complex environments.

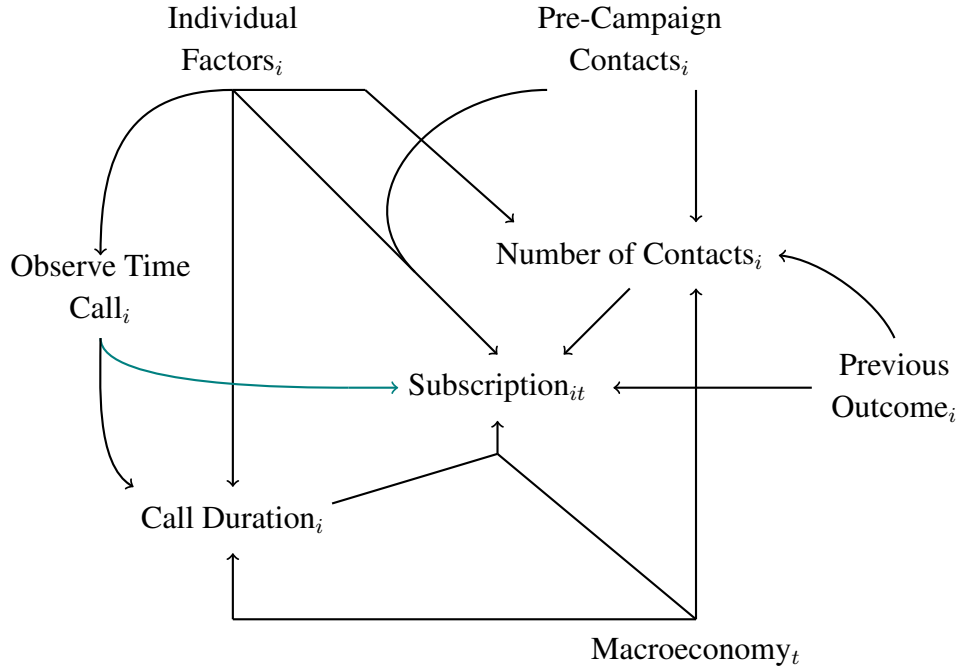
## 4 Call Timing Effects

We investigate the efficiency of call timing from both a within-week and within-month perspective.

### 4.1 DAG and Hypothesis

The directed acyclic graph (DAG) below illustrates the assumed causal relationships in a telemarketing campaign promoting term deposits. Our primary objective is to assess the impact of call timing—specifically, the day of the week and the time within the month—on the likelihood of subscription.

Figure 6: Relation Between Call-Time and Take-Up



The DAG distinguishes between two groups of exogenous variables: (i) individual-level socioeconomic characteristics such as occupation, marital status, and educational attainment, and (ii) macroeconomic indicators including employment rate, consumer prices, and confidence indices. These exogenous factors not only influence the outcome but also affect the timing and duration of the call. For instance, an individual's occupation may determine their availability to answer or return a call.

Call duration plays a critical causal role. Intuitively, very short calls are unlikely to result in product subscription, while longer calls tend to increase the probability of a successful

outcome. Duration is influenced by the timing of the call, as people may be more or less available depending on the time of day or the point in the month (e.g., due to work schedules or holidays). Duration is also affected by the number of calls made during the campaign and the aforementioned exogenous factors.

The DAG also incorporates variables from prior campaigns, such as the number of calls made and the outcomes achieved. These historical variables may influence both the current campaign’s call volume and its success rate.

## 4.2 Weekdays

Descriptive statistics suggested that mid-week may be a more effective time for outreach, possibly because individuals are more engaged in work-related activities and thus more open to financial conversations. However, selection effects may be present, as we only observe those who respond.

### 4.2.1 PICO Description

**Population:** Portuguese clients contacted by or contacting the bank

**Intervention:** Calls placed during the middle of the work week (assumed “business mindset”)

**Control:** Individuals with similar observable characteristics who receive calls at the beginning or end of the work week

**Outcome:** Probability of subscribing to the term deposit product

### 4.2.2 Results

Table 4: Change in take-up in middle of the week

Statistic	C-FOREST	MLPLR	MLPLR-ENS	MLIRM	LOGIT
Estimate	0.0125	0.0122	0.0087	0.0103	0.0151
Std Errors	0.0047	0.0047	0.0041	0.0048	0.0048
p-value	0.0079	0.0101	0.0314	0.0309	0.0015
CI 95%	0.0033-0.0217	0.0029-0.0215	0.0008-0.0167	0.0009-0.0196	0.0057-0.0244

All model specifications yield statistically significant results at the 5% level ( $p < 0.05$ ), with the largest marginal effects observed in the logit model. Estimates derived from the partial

linear model using an ensemble learner are the most conservative, likely due to its capacity to capture non-linear interactions. Similarly, the interactive regression model, which allows for fully heterogeneous effects, yields comparatively smaller estimates. This suggests that more flexible models tend to produce more attenuated effect sizes.

In terms of practical significance, a marginal effect of 0.012 implies a roughly 10% increase over the average take-up rate. This level of improvement may be of interest to financial institutions considering similar marketing strategies.

### 4.2.3 Heterogeneity Analysis

Given that occupation may moderate the effect of mid-week calls, we further analyze results across job categories.

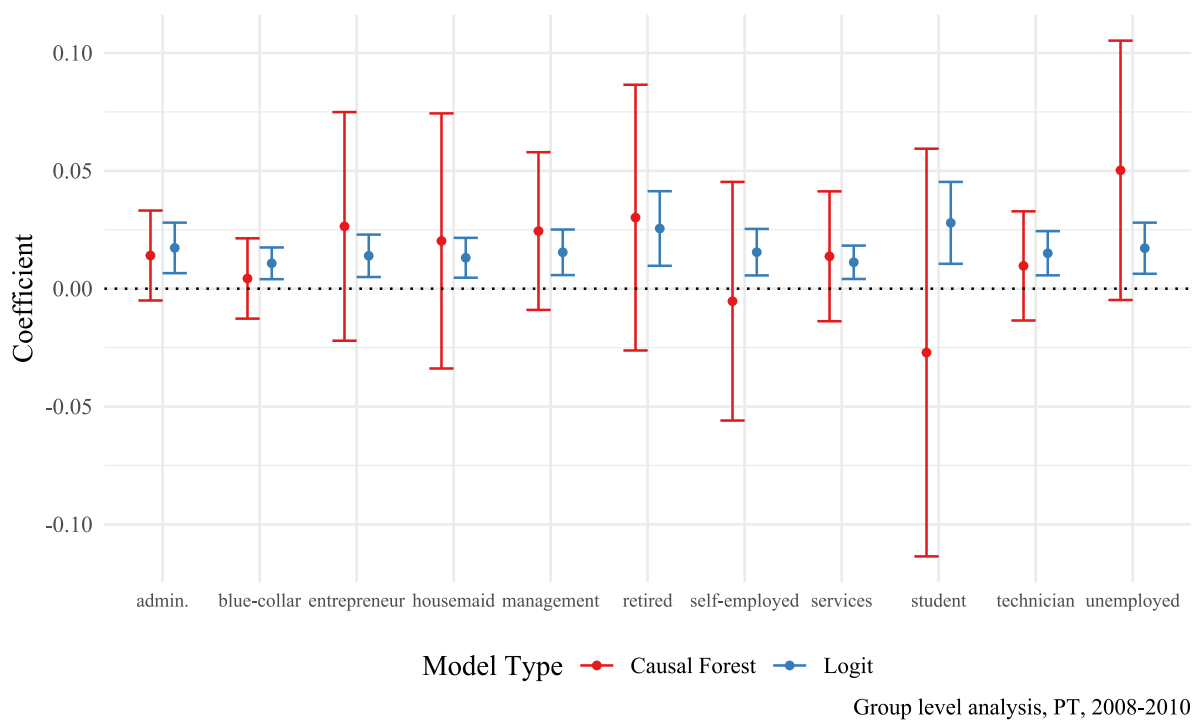


Figure 7: Change in take-up in middle of the week

Logit-based marginal effects are statistically significant across most subpopulations, with particularly strong results for administrative workers, retirees, and students. Estimates from the causal forest are more noisy as the model imposes less structure and our dataset is not particularly large, most results are not significant at the standard 5% level. However, effects for administrative

workers remain significant at the 5% level across both model types.

These findings indicate the presence of job-related heterogeneity in responsiveness to mid-week outreach.

### 4.3 Time of the month

We examine the effect of the time of the month on campaign success using a similar framework. A plausible hypothesis is that individuals may be more receptive to term deposit offers at the beginning of the month, when account balances are typically higher following salary payments.

#### 4.3.1 PICO Description

**Population:** Portuguese customers contacted by or contacting the bank

**Intervention:** Calls made in the second half of the month

**Control:** Individuals with similar observable characteristics who received calls in the first half of the month

**Outcome:** Probability of subscribing to the term deposit product

#### 4.3.2 Results

Table 5: Change in take-up end of month

Statistic	C-FOREST	MLPLR	MLPLR-ENS	MLIRM	LOGIT
Estimate	0.0117	0.0136	0.0079	0.0139	0.0148
Std Errors	0.0061	0.0060	0.0054	0.0061	0.0050
p-value	0.0533	0.0242	0.1447	0.0223	0.0032
CI 95%	-0.0002-0.0237	0.0018-0.0255	-0.0027-0.0186	0.0020-0.0259	0.0050-0.0246

Contrary to initial expectations, our results suggest that individuals are more likely to subscribe to a term deposit at the end of the month. These findings, however, appear less robust than those obtained for within-week variation. Once again, the logit model produces the largest estimated effects, whereas the partial linear model with an ensemble learner yields the most conservative estimates.

While these results should be interpreted with caution, they do provide some evidence that the timing of calls within the month can influence responsiveness to financial product offers.

### 4.3.3 Heterogeneity Analysis

We again extend the analysis by exploring heterogeneity across occupational groups. As in previous sections, results from the logit model remain consistently significant across groups, while the causal forest approach identifies statistically significant effects only for retirees. This suggests that retirees may be primarily responsible for the observed average treatment effect.

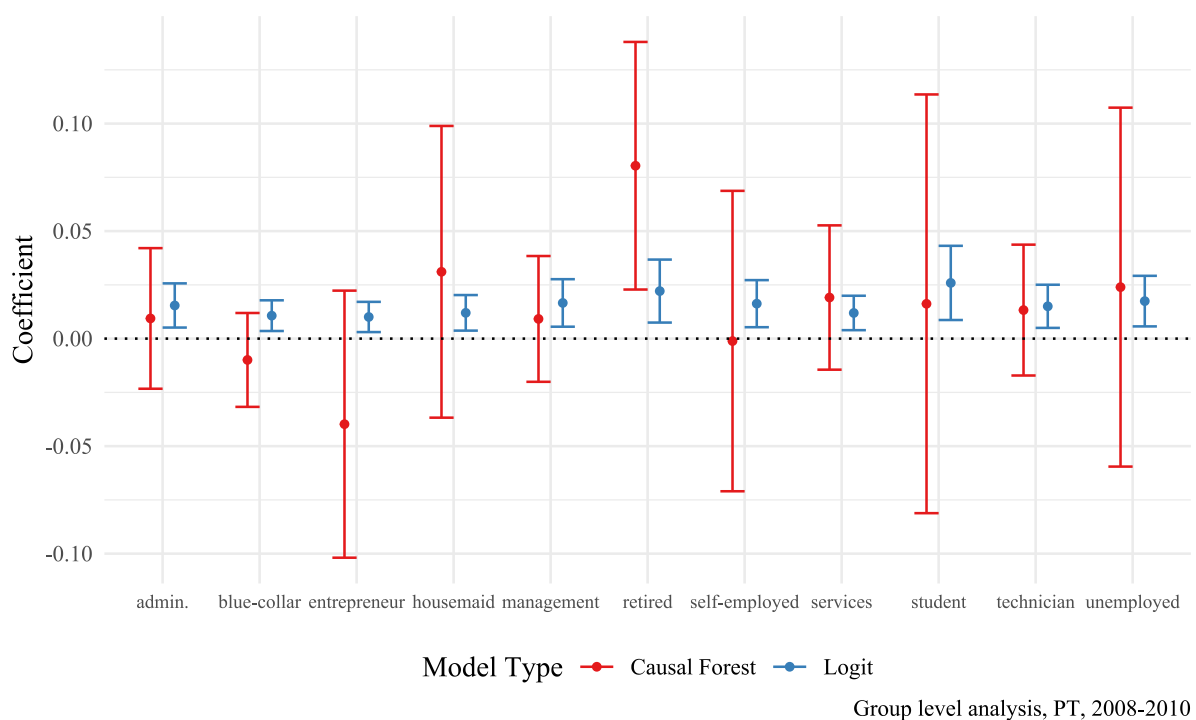


Figure 8: Change in take-up at the end of the month

We also find weak evidence of negative effects for entrepreneurs, who may be less inclined to commit to fixed-term products at the end of the month.



## 5 Macro Context

This section analyzes macroeconomic determinants of the marketing campaign, focusing on the impact of interest rates on product take-up rates. The specific rate offered to consumers is unobserved; however, the product, a term deposit, is among the fastest to react to changes in market conditions. The 12-month EURIBOR rate is used as a proxy, given its strong predictive power for such products. The objective is to estimate consumer demand elasticity with respect to interest rates. This has two key implications for the offering bank: it informs the expected effect of offering a more competitive rate or supports assessment of campaign viability when pricing power is limited and rates are determined by market conditions.

### 5.1 PICO summary

**Population:** The population of interest consists of bank consumers; however, due to previously mentioned data limitations, the analysis is restricted to clients who participated in the marketing campaign. To simplify the study design, only clients who received a single call are included.

**Intervention:** Variation in market interest rates which was particularly pronounced during the 2008–2010 period is leveraged to estimate demand elasticity.

**Comparison:** Clients exposed to lower market interest rates serve as a control group for those exposed to higher interest rates.

**Outcome:** The primary outcome of interest is the take-up of the offered savings product, measured as a change in the take-up rate.

### 5.2 DAG and Hypothesis

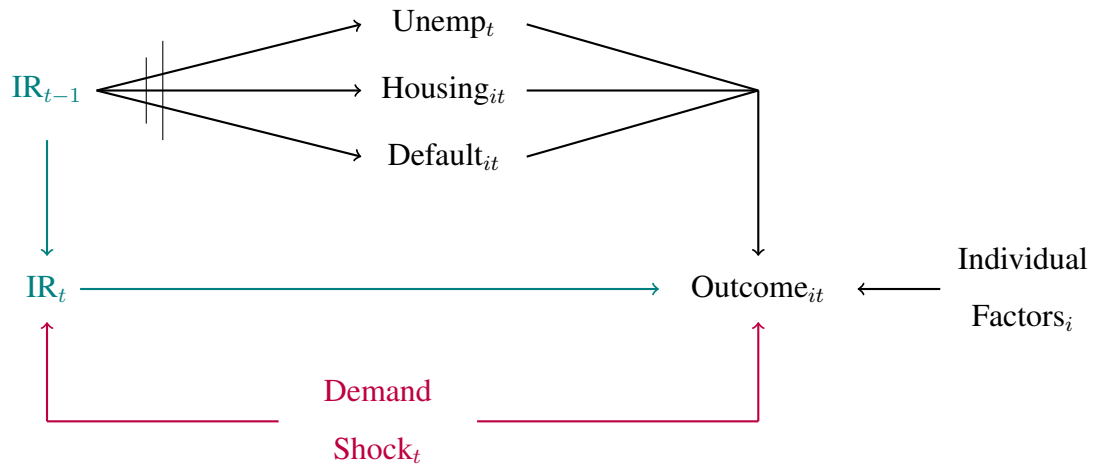
The main challenge in identifying the effect of interest rates on take-up rates is endogeneity. Specifically, interest rates represent market prices for capital and are thus the joint outcome of both supply and demand. For a reference on demand estimation see Berry and Haile 2021.

This is problematic because interest rates are correlated with demand shocks, which directly influence take-up behavior. A clear example is the bankruptcy of Lehman Brothers, which marked the onset of the great financial crisis.

This significantly altered investor risk preferences, leading to a substantial inflow of capital into safer financial instruments such as deposits or fixed term accounts.

The directed acyclic graph below summarizes the primary causal relationships assumed in this section.

Figure 9: Interest Rates and Take-Up



Under the structure outlined in the DAG, the lagged interest rate at period  $t - 1$  can be considered a valid instrument, provided that variables influenced by this rate are appropriately controlled for. The intuition behind this instrument is based on interest rate stickiness: financial products are often priced based on previously observed rates, while current market rates reflect present expectations rather than past ones. Therefore, contemporaneous demand shocks can be assumed to be uncorrelated with lagged interest rates.

This type of instrument and identification strategy is grounded in the macroeconomic literature (see Blundell and Bond 1998). However, the exogeneity of the instrument cannot be formally tested and could be questioned in our case. The absence of strong alternative instruments in our dataset rule out the possibility of conducting over-identification tests such as the Sargan–J test.

### 5.3 Results

To estimate causal effects, three methodological approaches are employed and compared: (i) a standard two-stage least squares (2SLS) estimator applied to a linear probability model, (ii) a double machine learning (DML) estimator for a partially linear instrumental variables (IV) regression, and (iii) a probit IV regression.

Table 6: Results from IV estimation

Statistic	LPM2SLS-IV	DML-IV	Probit-IV
Estimate	0.0473	0.0327	0.0185
Std Errors	0.0051	0.0028	0.0046
p-value	0.0000	0.0000	0.0001
Compute (s)	0.0577	39.5599	0.2065

All models yield statistically significant positive effects, indicating that higher interest rates increase the probability of subscription. The estimated local average treatment effects are economically substantial, ranging from 15% to nearly 40% of the baseline take-up rate.

The 2SLS linear probability model produces the largest point estimates, while the two-stage probit model yields the most conservative figures. The DML-IV estimator lies in between, likely capturing nonlinear effects missed by the LPM but without the parametric constraints of the probit model.

A remark which was not made often in this report is that compute differences are massive between traditional econometric methods and ml based approaches. Not accounting for tuning the DDML-IV was 791 times slower than the basic TSLS LPM model.

## **6 Conclusion**

### **6.1 Discussion**

This report provides empirical evidence that timing significantly influences the success of telemarketing campaigns. While traditional campaign strategies often focus on targeting the “right” individuals, our results suggest that “when” a potential client is contacted is equally important. For marketing professionals, this implies that optimal campaign design requires a dual focus: identifying both receptive individuals and favorable timing.

We employed a range of econometric and machine learning methods to estimate treatment effects and evaluate model performance. These approaches highlighted both the advantages and the limitations of flexible models in campaign evaluation and policy simulation.

Additionally, we estimated interest rate elasticities of demand for term deposit products, offering important insights for banks in their pricing strategies. These results underscore the relevance of incorporating demand-side behavior into macro-level financial planning.

### **6.2 Limitations**

Several limitations must be acknowledged. First, the data restricts the credibility and generalizability of our findings. Specifically, non-response bias may distort our estimates—individuals who do not answer phone calls may systematically differ in unobserved ways (e.g., being busier on certain days), potentially confounding our treatment effects.

Second, data constraints prevent analysis of follow-up call timing. Understanding when to re-engage with non-respondents is a critical component of effective campaign design. Unfortunately, we cannot address this question, but we encourage future work in this area.

Finally, our methodological comparisons would benefit from simulation studies under known and controlled data-generating processes (DGPs). Such studies would enable a more rigorous evaluation of estimator performance, particularly in contexts with limited sample sizes and complex interactions.

## 7 Bibliography

- Rasool Basha, Rasool Basha (Mar. 22, 2024). “A Study on the Effectiveness of Telemarketing in the Banking Industry”. In: *Shanlax International Journal of Management* 11 (S1-Mar), pp. 134–143. ISSN: 2581-9402, 2321-4643. DOI: 10.34293/management.v11iS1-Mar.8101.
- Berry, Steven and Philip Haile (Sept. 2021). *Foundations of Demand Estimation*. w29305. Cambridge, MA: National Bureau of Economic Research, w29305. DOI: 10.3386/w29305.
- Chernozhukov, Victor et al. (Feb. 1, 2018). “Double/debiased machine learning for treatment and structural parameters”. In: *The Econometrics Journal* 21.1, pp. C1–C68. ISSN: 1368-4221, 1368-423X. DOI: 10.1111/ectj.12097.
- Wager, Stefan and Susan Athey (July 3, 2018). “Estimation and Inference of Heterogeneous Treatment Effects using Random Forests”. In: *Journal of the American Statistical Association* 113.523, pp. 1228–1242. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.2017.1319839.
- Moro, Sérgio, Paulo Cortez, and Paulo Rita (June 2014). “A data-driven approach to predict the success of bank telemarketing”. In: *Decision Support Systems* 62, pp. 22–31. ISSN: 01679236. DOI: 10.1016/j.dss.2014.03.001.
- Ramey, Valerie A. (Feb. 2011). “Identifying Government Spending Shocks: It’s all in the Timing\*”. In: *The Quarterly Journal of Economics* 126.1, pp. 1–50. ISSN: 0033-5533, 1531-4650. DOI: 10.1093/qje/qjq008.
- Graham, John W. (Jan. 1, 2009). “Missing Data Analysis: Making It Work in the Real World”. In: *Annual Review of Psychology* 60.1, pp. 549–576. ISSN: 0066-4308, 1545-2085. DOI: 10.1146/annurev.psych.58.110405.085530.
- Sterne, J. A C et al. (Sept. 1, 2009). “Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls”. In: *BMJ* 338 (jun29 1), b2393–b2393. ISSN: 0959-8138, 1468-5833. DOI: 10.1136/bmj.b2393.
- Hawthorne, Graeme, Graeme Hawthorne, and Peter Elliott (July 2005). “Imputing Cross-Sectional Missing Data: Comparison of Common Techniques”. In: *Australian & New Zealand Journal of Psychiatry* 39.7, pp. 583–590. ISSN: 0004-8674, 1440-1614. DOI: 10.1080/j.1440-1614.2005.01630.x.
- Raghunathan, Trivellore E. (Apr. 1, 2004). “What Do We Do with Missing Data? Some Options for Analysis of Incomplete Data”. In: *Annual Review of Public Health* 25.1, pp. 99–117. ISSN: 0163-7525, 1545-2093. DOI: 10.1146/annurev.publhealth.25.102802.124410.

- Bennett, Derrick A. (Oct. 2001). “How can I deal with missing data in my study?” In: *Australian and New Zealand Journal of Public Health* 25.5, pp. 464–469. ISSN: 13260200. DOI: 10.1111/j.1467-842X.2001.tb00294.x.
- Blundell, Richard and Stephen Bond (Nov. 1998). “Initial conditions and moment restrictions in dynamic panel data models”. In: *Journal of Econometrics* 87.1, pp. 115–143. ISSN: 03044076. DOI: 10.1016/S0304-4076(98)00009-8.
- Rubin, Donald B. (June 9, 1987). *Multiple Imputation for Nonresponse in Surveys*. 1st ed. Wiley Series in Probability and Statistics. Wiley. ISBN: 978-0-471-08705-2 978-0-470-31669-6. DOI: 10.1002/9780470316696.

## A Appendix

### A.1 Additional statistics

We report some additional descriptive statistics.

Table 7: Observed Marital Status

Marital Status	Count	Share (%)
divorced	4612	11.20
married	24928	60.52
single	11568	28.09
unknown	80	0.19

Most of the individuals in our sample are married.

Table 8: Observed Education Level

Education	Count	Share (%)
basic.4y	4176	10.14
basic.6y	2292	5.56
basic.9y	6045	14.68
high.school	9515	23.10
illiterate	18	0.04
professional.course	5243	12.73
university.degree	12168	29.54
unknown	1731	4.20

We have a wide range of educational outcomes in our data and many professions are also represented in our sample.

Table 9: Observed Job Titles

Job Titles	Count	Share (%)
admin.	10422	25.30
blue-collar	9254	22.47
entrepreneur	1456	3.54
housemaid	1060	2.57
management	2924	7.10
retired	1720	4.18
self-employed	1421	3.45
services	3969	9.64
student	875	2.12
technician	6743	16.37
unemployed	1014	2.46
unknown	330	0.80