A Study on the Core Driving Factors and Heterogeneity of Customers' Subscription Behaviors for Fixed Deposits in Banks - An Empirical Analysis Based on Telephone Marketing Data from Portuguese Retail Banks

Zhao Ruohan
School of Economics and Management, Beijing Jiaotong University, Business Administration, Weihai, Shandong Province, 264200, China
emma01112026@163.com

ABSTRACT

Against the backdrop of increasingly fierce competition in the global retail banking industry, time deposits, as a stable source of bank liabilities, have become a key focus of precise marketing for commercial banks [1]. Based on the telephone marketing data of Portuguese retail banks from Kaggle, this study empirically explores the core driving factors and heterogeneity characteristics of customers' fixed deposit subscription behaviors from three dimensions. The results show that marketing historical success exerting the most prominent driving effect. In terms of heterogeneity, customers with different educational levels and loan status show significant differences in financial characteristics and subscription willingness. This study clarifies the internal logic of customers' fixed deposit subscription behaviors, and provides scientific and feasible strategic references for retail banks to optimize the allocation of telemarketing resources, implement stratified and precise marketing, and improve marketing conversion efficiency.

Keywords: Retail Banks; Subscription Behavior; Customer Heterogeneity; Precise Marketing

1. INTRODUCTION

1.1 Background

Nowadays, the global financial market competition has become increasingly fierce [8], and retail banking business has become the core engine for the profit growth of commercial banks, and time deposits serve as the cornerstone of a bank's stable liability sources [5], their business scale and customer structure directly determine the bank's funding costs and operational stability.

Empirical studies have confirmed that data-driven customer behavior analysis is the core solution to overcoming marketing inefficiencies in the financial industry [9]. Use empirical mining of large-scale marketing data, banks can systematically deconstruct the internal logic. At the same time, through heterogeneity analysis, they can accurately identify the decision-making differences among different customer groups. Therefore, the in-depth exploration of the core driving factors and heterogeneity characteristics of regular deposit subscription behavior is not only an important topic to respond to industrial needs, but also provides a typical sample for enriching customer behavior research in the field of financial marketing.

1.2 Research Questions

This research focuses on three core issues:

(1) Do these three core dimensions: customer financial status, marketing interaction quality, and historical marketing performance, have a significant impact on the subscription behavior of fixed deposits?

(2) Are there significant heterogeneities in the core driving mechanisms of subscription behavior among different customer groups?

(3) Based on the empirical conclusions, what is the feasible strategic references for the precision and efficiency of bank telemarketing?

2. LITERATURE REVIEW

2.1 Bank Customers' Subscription Behaviors for Financial Products

The behavior of bank customers subscribing to financial products is a classic research topic in the field of financial marketing. The existing research is relatively old and mainly analyzes this behavior from customer financial status and customer loyalty. For instance, the high-asset customers are less sensitive to interest rates [2]. And like customer loyalty, the depth of the customer' s cooperation with the bank and their past product holdings will strengthen their dependence and trust in the bank, thereby increasing the possibility of subsequent product subscriptions [6].

2.2 Customer Heterogeneity on the Effectiveness of Financial Marketing

Educational level is widely regarded as a key factor influencing customers' financial cognition and decision-making capabilities [7]. Customers with higher education levels have a deeper understanding of financial products and stronger ability to distinguish marketing information. Their subscription decisions are more dependent on rational analysis of product returns and risks [3]; while customers with lower educational levels are more susceptible to emotional factors such as marketing language and brand reputation [1]. Additionally, age, occupation, and other characteristics can also lead to differences in customers' risk preferences and product demands [4].

2.3 Deficiencies

Existing studies have comprehensively revealed the driving mechanisms behind bank customers' financial product subscription behaviors and preliminarily explored the moderating effect of customer heterogeneity. However, there are still limitations. Most of these studies employ multi-factor generalized analysis, incorporating numerous variables to construct complex models, which results in insufficient precision in identifying the core driving mechanisms. Secondly, heterogeneity analysis mostly focuses on simple grouping of demographic characteristics, and lacks in-depth exploration of the differences in intensity of core driving factors and the regulatory mechanisms among different groups. Finally, some research conclusions are overly theoretical and lack direct guidance for bank marketing practices, failing to provide practical and feasible marketing strategies for reference.

3. DESCRIPTIVE STATISTICS

3.1 Data Preprocessing

There are 775 blank values in the "balance" column of this dataset. After deleting these values, the sample size of the dataset becomes 10,388.

Meanwhile, the categorical variables are dummy-coded. The "tertiary" variable is coded as 1, and the others are coded as 0 to identify the heterogeneity impact of the high educational level group; the "With loan" variable is coded as 0, and "Without loan" are coded as 1 to distinguish the behavioral differences. Specifically, the historical marketing outcome variable is a three-level categorical variable encompassing "Failure/Unknown (0)", "Success (1)", and an intermediate reference category. To include it in the logistic model, it is decomposed into two dummy variables

3.2 Descriptive Statistics for Continuous Variables

Table 1. Descriptive Statistics of Continuous Variables

	Ave	Std.	Min	Median	Max
Balance	1642.43	3315.33	-6847	638	81204
Duration	373.60	348.19	2	257	3881

Table1 presents the descriptive statistics of two continuous variables, namely account balance and call duration.

3.3 Descriptive Statistics of Categorical Variables

Figure 1. Frequency distribution of core categorical variables related to deposit subscription

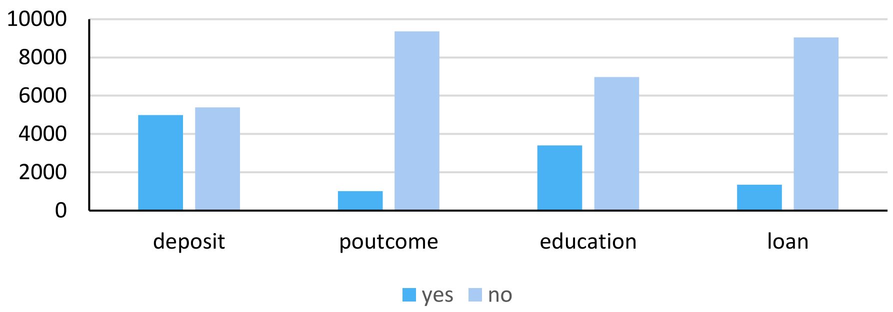

Figure1 intuitively presents customers who subscribed to “deposit”, had successful previous marketing outcomes, received higher “education”, and had no “loans” all significantly lower than their corresponding “no” counterparts, reflecting the sample’ s characteristic of a relatively small share of high-value subscribers and customers with favorable attributes.

3.4 Description of the Grouping Differences

3.4.1 Group Comparison of Interval Variables

Table 2. Group Comparison of Balance and Duration by Subscription Status

		t	df	P	Mean	Std.
Balance	Subscription Group (1)	7.934	10386	<0.001	1909.70	3573.90
	Unsubscribed Group (0)	7.886	9835.254	<0.001	1394.69	3035.58
Duration	Subscription Group (1)	50.958	10386	<0.001	535.34	392.59
	Unsubscribed Group (0)	49.895	7504.797	<0.001	223.69	209.85

Table2 presents the independent sample t-test results. For customer financial status, the average account balance of the subscription customers was significantly higher than that of the non-subscription customers. The Welch t-test showed that the difference was statistically significant ($p < 0.001$).

In terms of marketing interaction quality, the average call duration of the subscription customers was significantly longer than that of the non-subscription customers. The test also showed that the difference was statistically significant ($p < 0.001$).

These results preliminarily verified the positive correlation between variables, laying a descriptive foundation for the subsequent regression analysis.

3.4.2 Group Comparison of Categorical Variables

Table 3. Chi-square test results of group differences in categorical variables and subscription behavior

Variable grouping	Group level	Subscription percentage (%)	Chi-square value (χ^2)	df	p-value	Cramer's V coefficient
Outcome	Failure/Unknown (0)	43.4	850.601***	1	<0.001	0.275***
	Success (1)	91.2				
Education	Non-higher education (0)	44.8	93.851***	1	<0.001	0.095***
	Higher Education (1)	54.9				

Table 3. (continued) Chi-square test results of group differences in categorical variables and subscription behavior

Variable grouping	Group level	Subscription percentage (%)	Chi-square value (χ^2)	df	p-value	Cramer's V coefficient
Loan	With loan (0)	50.3	134.462***	1	<0.001	0.113***
	Without loan (1)	33.4				

Table3 shows that the differences in subscription behavior across the three groups are highly significant. The subscription rate of customers with a history of successful marketing is much higher (91.2%). Customers with higher education show a significantly higher subscription rate (54.9%) than those without (44.8%). Notably, customers with loans exhibit a higher subscription rate (50.3%) than those without loans (33.4%). The latter two variables show a weak but significant correlation with subscription behavior, providing a descriptive basis for subsequent logistic regression and heterogeneity analysis.

3.5 The Grouping Characteristics of the Heterogeneity Dimensions

In order to lays the groundwork for subsequent group regression analysis, fundamental differences among heterogeneous groups will be explored.

3.5.1 Educational Level

Table 4. Comparison of core variables among different educational level groups

GROUP	BALANCE MEAN (±SD)	DURATION MEAN (±SD)	SUBSCRIPTION RATE (%)	N
NON-HIGHER EDUCATION	1468.37 (±3093.43)	374.37 (±353.18)	44.8%	6982
HIGHER EDUCATION	1999.24 (±3704.16)	372.04 (±337.77)	54.9%	3406
TOTAL	1642.43 (±3315.33)	373.60 (±348.19)	48.1%	10388

Table4 shows the average account balance of higher education customers was significantly higher. The one-way ANOVA verified the statistical significance of the differences between the groups ($p < 0.05$). There was a negligible difference in the duration between the two groups. For subscription rate, it is significantly higher than that of non-higher education customers.

3.5.2 Loan Status

Table 5. Marketing Effectiveness Comparison by Loan Status

Group	Poutcome success rate	Duration mean Value	Subscription rate percentage
With loan	10.8%	373.82	50.3%
Without loan	3.6%	372.14	33.4%
Chi-square test	$p < 0.001$ (significant)	-	$p < 0.001$ (significant)

Table5 reveals significant differences ($p < 0.001$). Customers with loans show a higher subscription rate (50.3%) but lower past marketing success (3.6%), while those without loans have a lower baseline rate (33.4%) but much higher past success (10.8%).

4. EMPIRICAL TESTING AND MODEL VALIDATION

4.1 Multicollinearity and Homoscedasticity Test

To ensure the validity of the underlying assumptions of the regression model and to avoid erroneous results, some tests are conducted.

4.1.1 Test for Multicollinearity

Table6 shows the variance inflation factors of the core independent variables are all far below 2. This indicates that there is no serious multicollinearity problem among them.

Table 6. Variance Inflation Factor and Collinearity Diagnostics of Core Independent Variables

VARIABLE	VIF	TOLERANCE	CONCLUSION OF COLLINEARITY
BALANCE	1.002	0.998	No severe multicollinearity
DURATION	1.001	0.999	No severe multicollinearity
POUTCOME	1.002	0.998	No severe multicollinearity

4.1.2 Homogeneity of Variance Test

Table7 shows the Levene test for the balance and poutcome variables in the education group were conducted, and the result showed that the variances were not homogeneous ($p < 0.05$), which did not meet the prerequisites for ANOVA. Therefore, non-parametric tests or robust standard errors should be adopted.

Table 7. Homogeneity of variance test and descriptive statistics results

Variable	Dependent variable	Levene test p-value	Homogeneity of variance test	Group means (descriptions)
Education	Balance	<0.05	Uneven variance	Primary=1636.49, secondary=1387.44 Tertiary=1999.24, unknown=1850.88
Loan	Poutcome	<0.05	Uneven variance	No=0.11， yes=0.04

*Note: If the Levene test yields a p-value greater than 0.05, it indicates homogeneity of variances and parametric tests can be employed.

5. HYPOTHESIS TEST

In order to solve those three research questions, some hypothesis testing will be shown.

5.1 Main Effect Test of Core Driving Factors

Employing Logistic regression to test the impact of core independent variables on the subscription behavior of time deposits (coded as subscribe=1, not subscribe=0).

The Logistic regression model is defined as follows:

$$Logit\ P = \beta_0 + \beta_1 \times balance + \beta_2 \times duration + \beta_3 \times poutcome\ dummy + \varepsilon \tag{1}$$

Where:

$\boldsymbol{P}$ =Probability of customers subscribing to time deposits; $\boldsymbol{\beta_0}$ =Constant term of the model; $\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \boldsymbol{\beta_3} =$ Regression coefficients of balance, duration, and poutcome respectively; $\boldsymbol{\varepsilon} =$ Random error term

The test hypotheses are set as:

$$\boldsymbol{H_0}\text{: } \beta_1 = \beta_2 = \beta_3 = 0,\ \boldsymbol{H_a}\text{: at least one } \beta_i \neq 0 \tag{2}$$

$$\alpha = 0.05$$

Table 8. Binary Logistic Regression Results for Core Independent Variables

Variables	Coefficient	Std.Error	χ^2	Df	Sig.	Exp(B)	Lower	Upper
Balance	0.000	0.000	43.522	1	<0.001	1.000	1.000	1.000
Duration	0.005	0.000	1669.386	1	<0.001	1.005	1.004	1.005
Poutcome	2.926	0.000	644.406	1	<0.001	18.655	14.882	23.383
$\boldsymbol{\beta_0}$	-1.936	0.045	1860.512	1	<0.001	0.144	-	-

Table8 demonstrates that the overall model is statistically significant with a R^2 indicating it explains 41.7% of the variance in time deposit subscription behavior. Specifically, call duration has a significant positive effect; historical marketing outcome shows a strong positive impact, while account balance is statistically significant but economically negligible. And the regression equation is:

$$Logit\ P = -1.936 + 0.000 \times balance + 0.005 \times duration + 2.926 \times poutcome\ dummy + \varepsilon \tag{3}$$

5.2 Customer Heterogeneity Test

In order to answer the second research question, some steps will be shown below.

5.2.1 Analysis Based on the Heterogeneity of Educational Levels

Since the Levene test revealed that the variance of balance was not uniform among different education groups, the Welch test and Games-Howell post hoc test were adopted.

Set the hypothesis:

$\boldsymbol{\mu_1} =$ The average account balance of the clients in the higher education group

$\boldsymbol{\mu_2} =$ The average account balance of non-higher education group customers

$$\boldsymbol{H_0}\colon\ \mu_1 - \mu_2 = 0 \tag{4}$$

$$\boldsymbol{H_a}\colon\ \mu_1 - \mu_2 \neq 0 \tag{5}$$

$$\alpha = 0.05 \tag{6}$$

Table 9. Independent Samples Welch t-test Results for Account Balance by Educational Level

	Variable 1	Variable 2
Mean	1999.24	1468.37
Variance	13,720,842.91	9,569,309.16
Observations	3406	6982
Mean Difference	0.00	
Df	5570	
T Stat	7.68	
P(T<=t) one-tail	<0.001	
T Critical one-tail	1.645	
P(T<=t) two-tail	<0.001	
T Critical two-tail	1.96	

Table9 shows a statistically significant difference between two groups. The average monthly account balance in the higher education group is significantly higher than that in the non-higher education group.

$\boldsymbol{H_0}$ cannot be accepted. We can conclude that there is a significant difference in average account balances between different educational level groups.

Before conducting the heterogeneity analysis based on educational level, considering that the samples with negative account balances did not conform to the research context of client's disposable financial assets, this study conducted further screening of the original sample. Observations with a balance less than 0 were excluded, and the final valid sample size for the hypothesis verification was 9700.

To test the moderating effect of educational level on the relationship between balance and subscription behavior, this study constructed a Logistic regression model that included the independent variable, grouping variable, and interaction term. The model form is as follows:

$$Logit\ P = \beta_0 + \beta_1 \times balance + \beta_2 \times education\ dummy + \beta_3 \times balance\_edu + \varepsilon \tag{7}$$

education dummy: binary classification grouping variable

balance_edu: interaction term between account balance and the education grouping

Table 10. Results of Binary Logistic Regression with Interaction Term

	Regression coefficient	Std. Error	Wald value	Sig.	Exponentiated coefficient
Balance	0.000	0.000	2.951	0.086	1.000
Education dummy	0.363	0.076	23.003	<0.001	1.438
Balance_edu	0.000	0.000	0.062	0.803	1.000
Constant	-2.315	0.047	2414.185	<0.001	0.099

Table10 explains this assumption has not been verified. The positive impact of account balance on marketing subscription behavior does not show a stronger effect.

5.2.2 Heterogeneity Analysis Based on Loan Status

To examine whether the impact of historical marketing success on subscription behavior varies due to differences in individual loan status, the coefficient differences between the groups were tested by introducing the interaction term poutcome×loan.

Table 11. The test results of the regulatory effect of loan status

	B	OR	Sig.
Poutcome (1)	0.352	1.422	<0.001
Poutcome (2)	0.628	1.873	<0.001
Poutcome (3)	2.598	13.432	<0.001
Loan	0.655	1.926	<0.001
Poutcome_loan	0.206	1.229	0.653

Table11 shows good explanatory power: poutcome (3) (past success) yields an OR of 13.43, strongly boosting subscription likelihood, while loan (OR=1.93) indicates higher willingness among non-loan customers. The interaction term is insignificant (p=0.65), showing loan status does not alter the effect of prior success.

5.3 Probit Test

To verify the reliability of the research conclusions, this paper replaces the Logistic model with the Probit model and applies Winsorize. The Probit regression results show that the coefficient of call duration is always positive and highly significant at the 1% level, and has a stable and reliable positive driving effect on the subscription behavior of time deposits; the coefficient of balance is always negative and highly significant, and its inhibitory effect on the subscription behavior is not affected by extreme values; the historical marketing results are always not statistically significant in the Probit model, suggesting that its effect may depend on the model distribution assumption, but does not affect the core conclusion of this study.

6. RESEARCH CONCLUSIONS AND IMPLICATIONS

6.1 Conclusion

This study empirically examines core drivers of fixed deposit subscription behavior and analyzes heterogeneity by educational level and loan status. Three core variables show significant positive effects, with historical marketing performance being the most prominent. While account balance is statistically significant, its economic impact is negligible.

Educational level creates differences in financial traits and behavior but does not moderate the balance-subscription relationship, with no significant gap between high and non-high education groups. Loan status affects conversion efficiency and willingness, yet does not moderate the impact of historical performance; this success consistently boosts subscription probability for both loan and non-loan clients.

Overall, current bank telemarketing mainly targets ordinary clients, requiring improved exploration and reach for high-potential customers.

6.2 Theoretical Implications

The empirical results of this study provide new theoretical supplements and perspective expansions for the research on customer behavior in the field of financial marketing. It addresses the ambiguity of core mechanism identification in existing multi-factor generalized studies and provides scenario-specific empirical evidence for research on financial product subscription drivers. And the study establishes a replicable empirical framework for big data-driven telemarketing research. It adopts Welch' s test, Games-Howell post-hoc test, and robust standard errors to handle variance heterogeneity, and verifies moderating effects via Logistic regression with interaction terms. This provides standardized methodological references and strengthens the theoretical-practical linkage in financial marketing research.

6.3 Practical Insights

Based on the empirical findings, this study provides specific practical suggestions for retail banks to optimize their telephone marketing strategies for fixed deposits and enhance the efficiency and rationality of precise marketing and resource allocation.

6.3.1 Implement Precise Reach and Stratified Marketing

Prioritize non-loan customers as the core telemarketing group to leverage their higher conversion efficiency, while maintaining light, convenient contact with loan customers via digital channels. Tailor messaging: focus on rational details for educated customers, and emotional appeals with simplified explanations for others. Stratify outreach by past marketing success, prioritizing responsive leads and conducting targeted follow-ups for non-converters.

6.3.2 Optimize the Marketing Interaction Process

As marketing interaction quality drives subscription behavior, banks should refine communication duration control and train marketers. Pre-job training standardizes processes to boost information efficiency, while a professional system enhances need identification and emotion management. Simulated practices help adjust approaches, extending effective interaction to strengthen subscription intentions.

6.3.3 Optimize the Allocation Logic of Marketing Resources

They should expand telemarketing coverage to medium and low-asset customers to avoid missing potential subscribers, and design differentiated products, low-threshold flexible deposits for the mass market and customized large-denomination CDs for high-asset clients to boost subscription rates across groups.

6.4 Research Limitations and Future Prospects

6.4.1 Research limitations

First, data is solely from a Portuguese retail bank' s telemarketing from Kaggle, so the single geography and channel restrict external validity, limiting applicability to other regions, channels and scenarios. Second, it only examines heterogeneity in three variables, omitting key traits like age, occupation leading to incomplete analysis of subscription behavior heterogeneity and insufficient revelation of group decision differences. Third, it focuses on static customer traits and direct marketing factors, ignoring external variables and product features, requiring expansion of the model's explanatory dimensions.

6.4.2 Future Outlook

Further studies could be conducted in the following directions in the future.

It should first expand data sources to include cross-national banks and diverse marketing channels, conducting comparative studies to validate the current findings. Additionally, it should incorporate richer heterogeneity dimensions and online behavioral data, enabling a more comprehensive analysis of group-specific subscription mechanisms to support refined stratified marketing.

Studies should integrate macroeconomic factors like interest rates and product features into analytical models, examining their direct and interactive effects to enrich financial marketing behavior research. Finally, extending the research scope to wealth management, insurance and other financial products then test the generalizability of the identified mechanisms.

REFERENCE

[1] Alba, J. W. & Hutchinson, J. W. (1987). Dimensions of consumer expertise. Journal of Consumer Research, 13(4), 411–454.
[2] Argyle, B., Iverson, B. C., Kotter, J. D., Nadauld, T. & Palmer, C. (2026). The dynamics of retail deposit balances (NBER Working Paper No. w34742). National Bureau of Economic Research. Available at: https://www.nber.org/papers/w34742
[3] Chen, N. & Volpe, R. P. (1998). An analysis of personal financial literacy among college students. Financial Services Review, 7(2), 107–128.
[4] Grable, J. E. & Lytton, R. H. (1999). Financial risk tolerance revisited: the development of a risk assessment instrument. Financial Counseling and Planning, 8(3), 163-181.
[5] McKinsey & Company. (2026) The state of retail banking: Profitability and growth in the era of digital and AI. Available at: https://www.mckinsey.com/industries/financial-services/our-insights/the-state-of-retail-banking-profitability-and-growth-in-the-era-of-digital-and-ai
[6] Morgan, R. M. & Hunt, S. D. (1994). The commitment-trust theory of relationship marketing. Journal of Marketing, 58(3), 20–38.
[7] Lusardi, A. & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. Journal of Economic Literature, 52(1), 5–44.
[8] Tang et al (2022) Financial market resilience and financial development: A global perspective, Journal of International Financial Markets, Institutions and Money, Vol 80.
[9] Yerramsetty, T. (2025) AI-Enabled Financial Marketing: Leveraging ML, Predictive Analytics, and Data-Driven Strategies for Customer Engagement, SSRN Electronic Journal, 10(36S): 998-1008.