Banking Variable Rate Annuity Product: Variable Selection and Modeling Building

Alex Rodgers, Brett Laderman, Cameron Houston, Gina Barbagallo, Jason Wang

September 11, 2024

COMMERCIAL BANKING, CORP

Table of Contents

Banking Variable Rate Annuity Product: Variable Selection and Modeling Building

Overview

The Commercial Banking Corporation requested further analysis of the data.

The request entailed dealing with separation concerns, building logistic models, investigating possible interactions, and reporting final model variables ranked by significance.

We found that 14 of the 47 provided predictor variables were significant and useful for building a main effects model. Secondary accounts seemed to exhibit the strongest correlation with purchasing the product. Finally, we explored possible interactions with a forward selection model and found one additional notable interaction.

We cover the evaluation of the binned variables, highlight underlying connections, explore interactions, and provide detailed recommendations for next steps to create a more effective predictive model.

Methodology

This section covers the data used, missing values, and separation concerns in our analysis.

Data Used

Based on our previous report, we were provided a dataset that strategically binned all continuous variables to facilitate further analysis. We identified four variables in the dataset with missing values: the indicator for an investment account (INV), the indicator for a credit card (CC), the number of credit card purchases (CCPURC), and the indicator for home ownership (HMOWN). For each of these variables, we replaced the missing values with a "MISSING" category.

Separation Concerns

Our analysis identified separation concerns in two variables: the number of cash-back requests (CASHBK) and the number of money market credits (MMCRED). To account for this, all instances of CASHBK greater than 1, including both events and non-events, have been aggregated into a single category (1+). Similarly, all instances of MMCRED greater than 3 were aggregated into a single category (3+).

Analysis

This section covers model evaluation methods and the odds ratio of selected variables.

Model Evaluation

After addressing the missing values and adjusting for separation concerns, we identified the most effective model for the data. We started with a backward stepwise regression analysis, which found that 14 variables were optimal for our model. These variables are ranked in Table 1 by significance, with a cutoff threshold of 0.002.

Table 1: Backward Selection P Values

VARIABLE	P VALUE
Savings account balance (SAVBAL_BIN)	1.16E-128
Checking account balance (DDABAL_BIN)	7.73E-61
Certificate of deposit account balance (CDBAL_BIN)	3.40E-39
Money market account balance (MMBAL_BIN)	5.96E-23
Number of checks written (CHECKS_BIN)	1.74E-19
Total ATM withdrawal amount (ATMAMT_BIN)	9.03E-10
Number of teller visit interactions (TELLER_BIN)	1.96E-08
Indicator for credit card (CC)	1.43E-07
Indicator for checking account (DDA)	2.50E-05
Indicator for retirement account (IRA)	3.15E-05
Installment loan account balance (ILSBAL_BIN)	5.39E-05
Indicator for investment account (INV)	9.38E-05
Indicator for mortgage (MTG)	6.98E-04
Number of insufficient fund issues (NSF)	1.13E-03

Table 1 lists all variables that were significant at our defined cutoff value. Each significant variable is considered a potential predictor of the product purchase by clients. The lowest p-value indicates the highest significance, making savings account balance the most significant, with a p-value of 1.16E-128.

Using the 14 variables identified through backward selection, we explored potential interactions with forward selection. Table 2 ranks these variables from forward selection in order of significance based on p-value, again with a cutoff threshold of 0.002.

Table 2: Forward Selection P Values

VARIABLE	P VALUE
Savings account balance (SAVBAL_BIN)	8.01E-129
Checking account balance (DDABAL_BIN)	5.63E-60
Certificate of deposit account balance (CDBAL_BIN)	2.58E-39
Money market account balance (MMBAL_BIN)	2.37E-23
Number of checks written (CHECKS_BIN)	5.76E-20
Total ATM withdrawal amount (ATMAMT_BIN)	9.09E-10
Number of teller visit interactions (TELLER_BIN)	1.93E-08
Indicator for credit card (CC)	1.53E-07
Indicator for checking account (DDA)	1.12E-05
Installment loan account balance (ILSBAL_BIN)	5.64E-05
Indicator for investment account (INV)	1.15E-04
Interaction between indicator for retirement account and checking account (IRA: DDA)	3.13E-04
Indicator for mortgage (MTG)	6.62E-04
Number of insufficient fund issues (NSF)	9.57E-04
Indicator for retirement account (IRA)	8.80E-01

In the final forward selection model, we identified one significant interaction between indicators for retirement accounts (IRA) and checking accounts (DDA). Table 2 lists variables found significant at our defined cutoff value, including the main terms that were not significant (IRA). Each of these variables is used in our predictive model of client product purchases.

Odds Ratio

We were interested in looking at the variable with the highest significance, savings account balance. We found that the odds of product purchase generally increased across higher savings account balance bins. Each balance bin below is compared to the base bin of less than or equal to $0.01 and is written with the odds of purchasing:

- Bin *$0.01 < balance ≤ $61.25* is 1.76 times less likely
- Bin *$61.25 < balance ≤ $265.87* is 1.29 times less likely
- Bin *$265.87 < balance ≤ $1259.45* is 1.35 times more likely
- Bin *$1259.45 < balance ≤ $2962.02* is 2.43 times more likely
- Bin *$296.02 < balance ≤ $8334.97* is 3.71 times more likely
- Bin *balance ≥ $8334.97* is 5.97 times more likely

This finding suggests that product cost and amount of disposable income are likely factors that customers weigh heavily when deciding whether to purchase an annuity. The result is somewhat counterintuitive however because when large amounts of money are kept in a savings account that is being done at the expense of higher yield accounts, exchanging high yield for low-to-zero risk. Variable rate annuities have a higher risk component that savings accounts lack, so the fact that customers loading large sums of money into a low-risk account like a savings account are also purchasing variable rate annuities is surprising. It could be that the ability to afford a variable rate annuity is a more important factor in the purchase of one than a customer's intentions when they joined The Bank, and that customers with large savings accounts represent an untapped vein of people who would make more strategic investments if educated about them.

Results

Our analysis of multiple models found that all variables listed in Table 2 were significant predictors of product purchase, with savings account balance being the most significant.

We recommend directing product annuity marketing efforts towards individuals with larger balances in secondary accounts, such as savings accounts, certificates of deposit, and money market accounts, as these account balances have shown a stronger correlation with purchasing the product. Our analysis revealed savings account balance as the most significant predictor; with the highest bin being $8334.97 or greater, we found those clients are 5.9 times more likely to purchase the insurance compared to those in lower balance bins.

Focusing on customers with more significant balances in these secondary accounts could enhance the effectiveness of your marketing strategies and improve purchase rates.

Conclusion

After being tasked with model creation to predict client product purchases, we addressed separation concerns, built logistic models, investigated possible interactions, and reported final model variables by significance. We found 14 statistically significant variables in relation to the purchase of the annuity product.

Secondary accounts exhibited the strongest correlation with purchasing the product, with savings account balance being the most significant predictor overall. There was one positive interaction between possessing both a retirement account and a checking account that was significant in predicting whether the product was purchased.

Appendix

Appendix Table 1: Correlation coefficients between product purchase (INS) and other variables

Variable	Estimates
INS	1.00
SAVBAL_BIN	0.24
CDBAL_BIN	0.21
CD	0.20
MMBAL_BIN	0.17
MM	0.17
CC	0.15
SAV	0.14
IRA	0.14
IRABAL_BIN	0.13
CCBAL_BIN	0.11
INV	0.11
CCPURC	0.11
MMCRED	0.09
INVBAL_BIN	0.09
SDB	0.07
POSAMT_BIN	0.06
DDABAL_BIN	0.05
BRANCH	0.05
HMVAL_BIN	0.04
MTGBAL_BIN	0.01
ACCTAGE_BIN	0.01

Variable	Estimates
LOC	0.01
MTG	0.01
INCOME_BIN	0.00
TELLER_BIN	0.00
LORES_BIN	-0.01
HMOWN	-0.01
CRSCORE_BIN	-0.01
LOCBAL_BIN	-0.01
AGE_BIN	-0.01
MOVED	-0.01
RES	-0.01
POS_BIN	-0.02
PHONE_BIN	-0.03
ILS	-0.03
ILSBAL_BIN	-0.03
CASHBK	-0.04
DEPAMT_BIN	-0.05
INAREA	-0.05
ATMAMT_BIN	-0.07
NSF	-0.07
DIRDEP	-0.07
NSFAMT_BIN	-0.07
CHECKS_BIN	-0.10
ATM	-0.12

Variable	Estimates
DEP_BIN	-0.16
DDA	-0.19