Banking Variable Rate Annuity Product: Variable Understanding and Assumptions

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

August 30, 2024

COMMERCIAL BANKING, CORP

Table of Contents

BANKING VARIABLE RATE ANNUITY PRODUCT: VARIABLE UNDERSTANDING AND ASSUMPTIONS

Overview

You reached out to explore ways to predict your clients' likelihood of purchasing a variable rate annuity. Our team began finding ways in which you could do so by analyzing your dataset; the data included **47** potential predictor variables related to client attributes and behaviors to help identify potential predicting characteristics.

Our analysis revealed that over 35% of the variables were not useful for predicting purchases based on significance alone, and among the 13 continuous significant variables, 12 did not meet the necessary assumptions.

In this report, we will evaluate the variables, highlight underlying insights not immediately seen in testing, and outline our recommended next steps to help you create an effective predictive model.

Methodology

Data Used

The initial analysis phase for modeling the product purchases at the Bank aimed to evaluate the significance of various predictor variables relative to the target variable, which moving forward, will be variable rate annuity purchase (**INS**).

The target variable is binary and uses 0 to indicate when a client did not/has not purchased the variable rate annuity, and 1 indicates that they did purchase the product. Next, all predictor variables were categorized by continuous, nominal, ordinal, or binary types.

Significant Variables

Each potential predictor variable was then tested for significance; the following tests were used:

- Pearson's chi-square test for nominal variables.
- Mantel-Haenszel chi-square test for ordinal and binary variables.
- Binary logistic regression for continuous predictor variables.

We established a significance threshold of **0.002 (0.2%)**. This criterion enabled us to identify **29** predictor variables with significant p-values in relation to **INS**. Table 1 presents these variables ranked by significance.

Table 1: Significant Predictors of INS

VARIABLES	CLASS	P VALUE
Indicator for certificate of deposit account	Binary	2.97E-78
Indicator for checking account	Binary	5.50E-70
Indicator for money market	Binary	7.45E-57
Saving account balance	Continuous	7.82E-48
Money market balance	Continuous	3.58E-47
Indicator for savings account	Binary	1.44E-39
Checking deposits	Continuous	2.31E-38
Indicator for retirement account	Binary	4.89E-37
CD balance	Continuous	1.05E-36
Indicator for credit card	Binary	2.44E-32
Indicator for ATM transactions	Binary	1.59E-29
Indicator for investment account	Binary	1.05E-21
Checking account balance	Continuous	1.99E-23
Number of telephone banking interactions	Continuous	1.43E-19
Number of money market credits	Ordinal	8.47E-16
Branch of bank	Nominal	5.49E-14
Number of insufficient fund issues	Ordinal	4.93E-11
Indicator for direct deposit	Binary	3.67E-11
IRA balance	Continuous	3.27E-11
Number of checks written	Continuous	1.92E-11
Value of home	Continuous	6.66E-12
Number of credit card purchases	Ordinal	1.07E-10
Indicator for safety deposit box	Binary	4.31E-10
Total ATM withdrawal amount	Continuous	5.85E-08
Indicator for local address	Binary	7.45E-07
Number of point of sale interactions	Continuous	6.22E-07
Number of cash back requests	Ordinal	7.06E-04
Amount of NSF	Continuous	1.35E-04
Total amount deposited	Continuous	3.72E-04

Table 1 lists all variables that we found significant with our defined cut-off value. Each variable that is significant is considered as potential predictors of the product purchase by clients. In the table, the lowest p-value indicates the highest significance.

Analysis

Odds Ratio

Binary predictor variables significant at the **0.002 (0.2%)** alpha level were used to create the following table of odds ratios in relation to **INS**. The interpretations for variables are as follows:

Individuals with investment accounts **(INV)** are ***3.47*** times as likely to purchase the variable rate annuity.

Table 2: Odds ratios for significant binary variables

VARIABLES	CLASS	ODDS RATIO
Indicator for investment account	Binary	3.47
Indicator for certificate of deposit account	Binary	3.43
Indicator for retirement account	Binary	3.18
Indicator for money market account	Binary	2.85
Indicator for savings account	Binary	1.83
Indicator for credit card	Binary	1.78
Indicator for safety deposit box	Binary	1.55
Indicator for direct deposit	Binary	0.71
Indicator for ATM interaction	Binary	0.59
Indicator for local address	Binary	0.57
Indicator for checking account	Binary	0.38

Table 2 highlights differences in odds ratios for certain predictors. The largest, **INV**, increases odds by 247%, whereas **DDA** decreases the odds by 62% This highlights the impact that certain variables have on the odds of purchase.

Medium-to-long-term investments seem to affect a customer's likelihood to purchase a variable rate annuity. The three variables associated with the strongest increase in odds of a variable rate annuity purchase, **INV**, Certificate of Deposit (**CD**), and Retirement Accounts (**IRA**), are all geared towards a future income goal.

Assumptions

We tested all significant continuous variables for linearity assumptions and found that only one variable, the value of one's home (**HMVAL**), met both significance and linearity cutoffs. The other significant continuous variables—**DDABAL, DEP, DEPAMT, CHECKS, NSFAMT, PHONE, SAVBAL, ATMAMT, POS, CDBAL, IRSBAL,** and **MMBAL**—did not meet the linearity assumption.

Additional Variable Information

In our analysis, we found 15 out of 48 variables with missing values, six of which were significant in predicting the purchase of variable rate annuities. As Figure 1 shows, **AGE** has the largest amount of missing entries at 1,702. However, amongst the significant variables, **HMVAL** had the most with 1,537 entries missing.

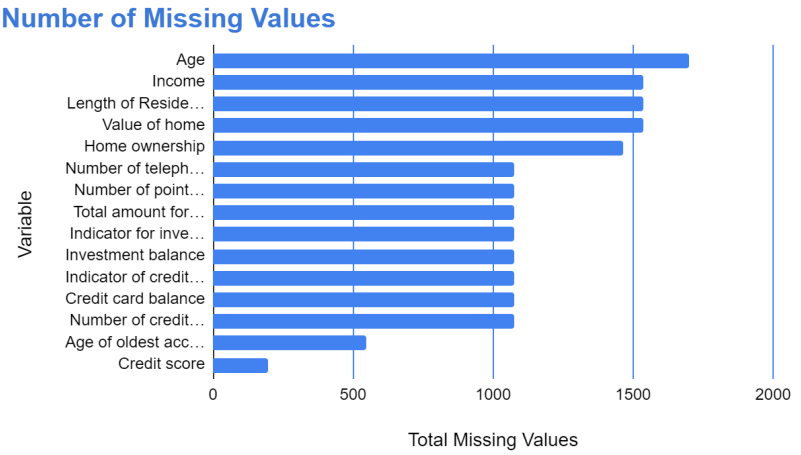

Figure 1: Missing Values by Variable

Figure 1 shows variables with missing values in the data, ranked from most to least.

We found that some variables show a redundancy of information. Specifically, all the balance variables and their respective indicators can be seen as redundant information collected by the Bank in cases where there is not a zero balance.

Interesting Findings

Having a checking account was associated with lower odds of variable rate annuity purchase while the possession of medium-to-long-term investment accounts such as IRAs and CDs was associated with higher odds of variable rate annuity purchase. This could indicate that the type of customer that is likely to invest in a variable rate annuity rate shops each type of account. They could have come to the Bank due to more competitive rates on investment-type accounts while using a different bank with more favorable rates for checking accounts for their daily use.

Recommendations

Owners of CDs, investment accounts, and IRAs have higher odds of purchasing the variable rate annuity, we suggest allocating a larger share of variable rate annuity marketing to this group. It will likely be worthwhile to have a predictive model built from the variables we found to be statistically significant in order to more finely tune your marketing beyond just those who already have medium-to-long-term investment accounts. Several variables had quite a few missing data points, many of which should be standard to obtain when a customer joins the Bank, such as age, or when a customer opens certain types of accounts, such as income. Because models are most effective when there is little to no missing data, we recommend looking into the cause of certain information not getting captured. This will allow for future predictive models to be even stronger.

Conclusion

Out of the 47 original variables, only 29 were statistically significant in relation to the purchase of the annuity product. With **INV**, **CD**, and **IRA** variables being associated with the largest increase in odds of purchase and the **ATM**, **DDA** variables being associated with the largest decreases in odds of purchase. Of the statistically significant continuous variables only one, **HMVAL** met linearity assumptions.

Appendix

ATTRIBUTES	TYPE	P VALUE
CD	Binary	2.97E-78
DDA	Binary	5.50E-70
MM	Binary	7.45E-57
SAVBAL	Continuous	7.82E-48
MMBAL	Continuous	3.58E-47
SAV	Binary	1.44E-39
DEP	Continuous	2.31E-38
IRA	Binary	4.89E-37
CDBAL	Continuous	1.05E-36
CC	Binary	2.44E-32
ATM	Binary	1.59E-29
INV	Binary	1.05E-21
DDABAL	Continuous	1.99E-23
PHONE	Continuous	1.43E-19
MMCRED	Ordinal	8.47E-16
BRANCH	Nominal	5.49E-14
NSF	Ordinal	4.93E-11
DIRDEP	Binary	3.67E-11
IRABAL	Continuous	3.27E-11
CHECKS	Continuous	1.92E-11
HMVAL	Continuous	6.66E-12
CCPURC	Ordinal	1.07E-10
SDB	Binary	4.31E-10
ATMAMT	Continuous	5.85E-08
INAREA	Binary	7.45E-07
POS	Continuous	6.22E-07
CASHBK	Ordinal	7.06E-04
NSFAMT	Continuous	1.35E-04
DEPAMT	Continuous	3.72E-04
CCBAL	Continuous	3.22E-03
ILS	Binary	7.27E-03
ACCTAGE	Continuous	7.87E-03
TELLER	Continuous	9.31E-03
ILSBAL	Continuous	3.13E-02

INVBAL	Continuous	3.90E-02
MTGBAL	Continuous	5.95E-02
POSAMT	Continuous	1.19E-01
AGE	Continuous	2.19E-01
RES	Nominal	2.34E-01
MOVED	Binary	2.37E-01
INCOME	Continuous	2.57E-01
CRSCORE	Continuous	3.93E-01
LOC	Binary	4.99E-01
MTG	Binary	5.28E-01
LORES	Continuous	8.51E-01
LOCBAL	Continuous	9.11E-01
HMOWN	Binary	9.20E-01

var1 <chr>	Class <chr>	p_values <dbl>	linearity <chr>	linearity_p_val <dbl>
DDABAL	Continuous	1.994710e-23	failed	1.583784e-13
DEP	Continuous	2.312173e-38	failed	5.563859e-12
DEPAMT	Continuous	3.723108e-04	failed	9.129117e-17
CHECKS	Continuous	1.924075e-11	failed	1.488086e-05
NSFAMT	Continuous	1.347531e-04	failed	8.435916e-06
PHONE	Continuous	1.431599e-19	failed	2.116996e-05
SAVBAL	Continuous	7.821331e-48	failed	7.123709e-100
ATMAMT	Continuous	5.847175e-08	failed	2.049616e-34
POS	Continuous	6.217900e-07	failed	3.278489e-10
CDBAL	Continuous	1.052622e-36	failed	4.122965e-22
IRABAL	Continuous	3.273501e-11	failed	1.283673e-10
MMBAL	Continuous	3.576739e-47	failed	9.748418e-06
HMVAL	Continuous	6.663097e-12	passed	1.269931e-02