Complex Objective functions:
Predicting Large Deposits of Wealth Management Clients

1. Background

In practical applications of both developing and applying models, the use of statistics and predictive models has become quite common in many business applications. What was normally a technique confined to large direct mailers has now expanded to virtually all media and as a consequence to virtually all business sectors. Certainly, the notion of developing models to predict a specific behaviour and the ability to demonstrate the tangible business benefits of doing so is not a new phenomenon. But what if the desired or predicted behaviour is not necessarily straightforward? In developing predictive models, one of the preliminary tasks is the creation of the objective function i.e. the behaviour we are trying to predict.

In developing response models or defection models, creating the objective function can be as simple as determining whether or not a given person purchased an item (response) or whether or not a person has become inactive (defection). Once the objective function is determined, the modeling process attempts to determine the characteristics that best predict the behaviour in question. Straightforward with respect to response or defection, this process can become quite complicated if our objective function is not well defined.

What follows provides description of a solution to a modeling challenge arising out of a Client need to identify a difficult-to-define behaviour.

A Boire Filler Group Client manages the investment portfolios of high-income wealth management clients. Initially, the Client desired to produce a solution that identified clients within its portfolio that were at risk of moving their entire investment portfolio to another manager. In this case, the objective function could be determined by identifying those clients that redeemed their entire portfolio. From there, it was simply a case of identifying the key triggers that predict complete redemption behaviour.

In addition to seeking the capability to preemptively identify negative behaviour such as risk of defection, the Client also wanted to preemptively identify an important positive behaviour – predict those clients that are more likely to make very large deposits. A key challenge in developing this large deposit model was to determine how to define the objective function of predicting large deposits.

Does a deposit of $100k constitute a large deposit if the investor historically has made average deposits of this size? By conducting certain analytics exercises, we were able to determine the appropriate rules to define and identify large deposit behaviour. With the objective function created, our standard modeling

process was applied in achieving a solution that predicted the likelihood of a high income wealth management client making a large deposit.

2. The Approach

As we attempted to frame the problem, we realized that the behaviour of identifying large deposits does not necessarily relate to an absolute deposit amount. Rather, a large deposit may be determined based on how an individual's deposit *compares* to past deposit behaviour. In comparing current deposit behavior to previous deposit behaviour, we are attempting to discern whether or not this most recent deposit behaviour is out of pattern with the historical deposit behaviour. However, just simply looking at how deposit behaviour has changed between the recent period and the past may not be indicative of large deposit behavior. From our Client's perspective, a 300% increase in deposit does not necessarily constitute a large deposit if the client made a deposit of $300 even though his average deposit overtime has only been $100. So in defining large deposit behaviour, we looked at two key criteria:

- Percent *change* in deposit behaviour
- Total Deposit *Amount*

Having identified the above two metrics as the key pieces of information required to define large deposit behaviour, we then needed to determine the appropriate client base from which the model would be built. Two scenarios were examined - the first being clients with a 50% change in deposit behaviour, and the second scenario being clients that had a 25% change in deposit behaviour. The 50% group included 18,439 high income clients. We then produced three distributions within this group. Each distribution represented a matrix with *% change in deposit* on the y-axis (here 50% change in deposit was the minimum) and *deposit amount* on the x-axis. The % change numbers are grouped in deciles, meaning that the file of clients are sorted in descending order by size of % change, with clients in decile 1 having the highest % change and clients in decile 10 having the lowest percent change. The first report looks at the % change in deposit between the most recent quarter and the quarter prior to most recent:

	NO. OF CLIENTS					
RANK_PERCENT_CHANGE	< 25K	25K - 50K	50K - 75K	75K - 100K	100K+	Grand Total
1	403	533	240	134	520	1830
2	1868					1868
3	1862				1	1863
4	1212	171	113	49	291	1836
5	1063	312	185	81	207	1848
6	1318	276	115	39	101	1849
7	1472	186	92	44	55	1849
8	1544	185	65	19	37	1850
9	1651	104	55	14	24	1848
10	1611	103	40	16	28	1798
Grand Total	14004	1870	905	396	1264	18439

In this initial report clients in the small deposit amount (<$25) group (highlighted in yellow) display no trend when ranked by percent change. This indicates that % change and deposit amount distribution are essentially unrelated at low deposit amounts. In fact, for very high percent change values, fewer clients have low deposits. But, as deposit amount increases we see the trend reversing; there is a much clearer trend between high % deposit *change* and high deposit *amount*, with the best trend being at $100k+ deposit amount. However, one significant peculiarity with this report is the fact that there are virtually zero clients in % change deciles 2 and 3 with deposit amounts greater than $25K. When looking at this data, we surmised that the reason for this peculiarity might be due to instability when measuring % change over a short period of time. As an alternative to resolve this issue we decided to look at the most recent quarter compared to the *average of the last 3 quarters* to determine if we could obtain more stability in the numbers, especially in % change deciles 2 and 3 with deposit amounts >$25K. The table below shows results of this revised approach:

	NO. OF CLIENTS					
RANK_PERCENT_CHANGE	< 25K	25K - 50K	50K - 75K	75K - 100K	100K+	Grand Total
1	403	533	240	134	520	1830
2	1858					1858
3	1868				1	1869
4	1203	169	111	48	288	1819
5	1058	312	186	82	206	1844
6	1315	276	114	39	100	1844
7	1464	188	93	42	57	1844
8	1542	183	62	20	37	1844
9	1649	103	55	13	24	1844
10	1658	100	41	16	28	1843
Grand Total	14018	1864	902	394	1261	18439

Based on the above result, we were unable to resolve the problem of instability in percent change deciles 2 and 3. Our next step was to look at defining % change in deposit as the change in deposit between the most recent quarter and *all previous quarters*. Listed below are the results of this 3rd analysis:

	NO. OF CLIENTS					
RANK_PERCENT_CHANGE	< 25K	25K - 50K	50K - 75K	75K - 100K	100K+	Grand Total
1	699	377	186	104	477	1843
2	1414	92	53	38	248	1845
3	561	292	206	123	662	1844
4	722	383	252	131	348	1836
5	951	395	189	89	228	1852
6	1114	317	139	66	165	1801
7	1283	299	137	52	116	1887
8	1275	252	132	76	101	1836
9	1457	226	77	33	59	1852
10	1625	118	49	16	35	1843
Grand Total	11101	2751	1420	728	2439	18439

Under this scenario, we see stability in % change deciles 2 and 3 as deposit amount increases. In fact at the $100k+ deposit amount, we see that the distribution skews toward the larger percent change deciles (deciles 1 to 3).

We did look at similar matrixes under the scenario of a 25% change in deposit between periods but our trend results at the high deposit amount level were much less pronounced.

With these results, we were confident that clients exhibiting a % change in deposit amount of >50% with an *absolute* deposit amount of 100+ were exhibiting out of pattern behaviour that could be identified as a large deposit amount.

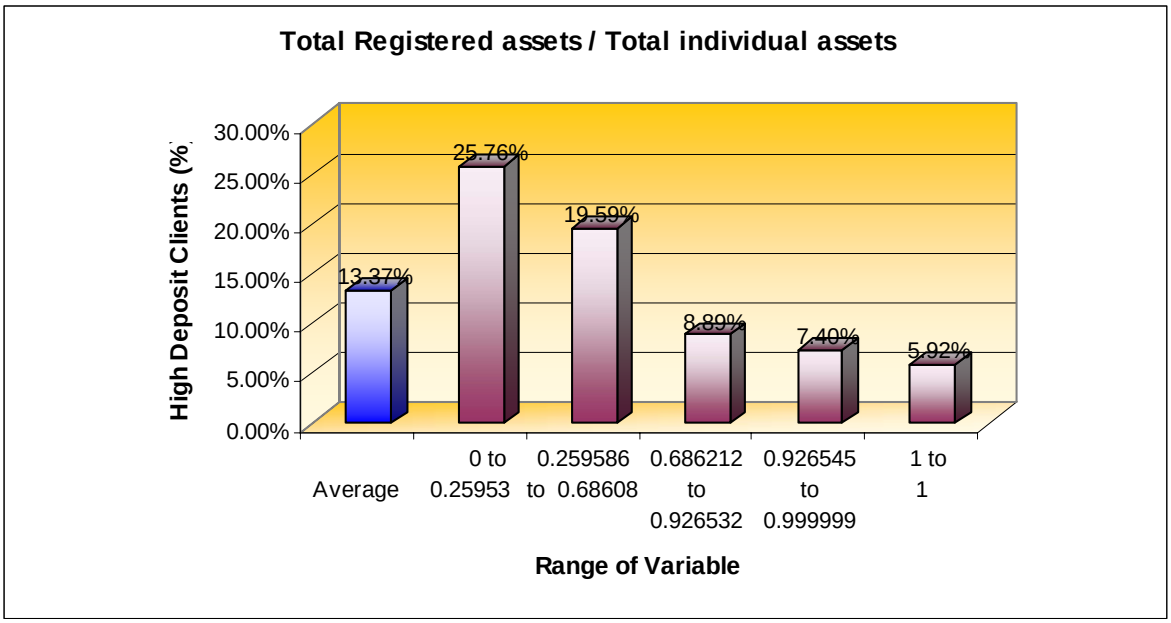
3. Modelling Methodology

Having defined the objective function of large deposit behaviour in our sample of 18,439 records at the customer or client level, we then applied our standard modeling methodology to this exercise.

The process of reviewing all data to better understand its meaning and relevance, as well as identifying any data integrity issues was addressed through our standard data audit routines. At the end of these routines we are able to determine how to build the independent variables for this model. It was determined that 458 variables could be created. The analytical file of 18,439 clients was then split 50/50 into the development sample and the holdout or validation sample. The objective function of large deposit rate was approximately 13% for both sample files. In constructing this analytical file, the objective function was defined in our post period. All independent variables are then created and developed in periods prior to (pre-period) the post period.

Correlation analysis reduced the 458 variables to 167 variables. EDA or exploratory data analysis was conducted on each of these remaining statistically significant variables to better understand the relationship of each variable with large deposit behaviour.

An example of one such variable is listed below.



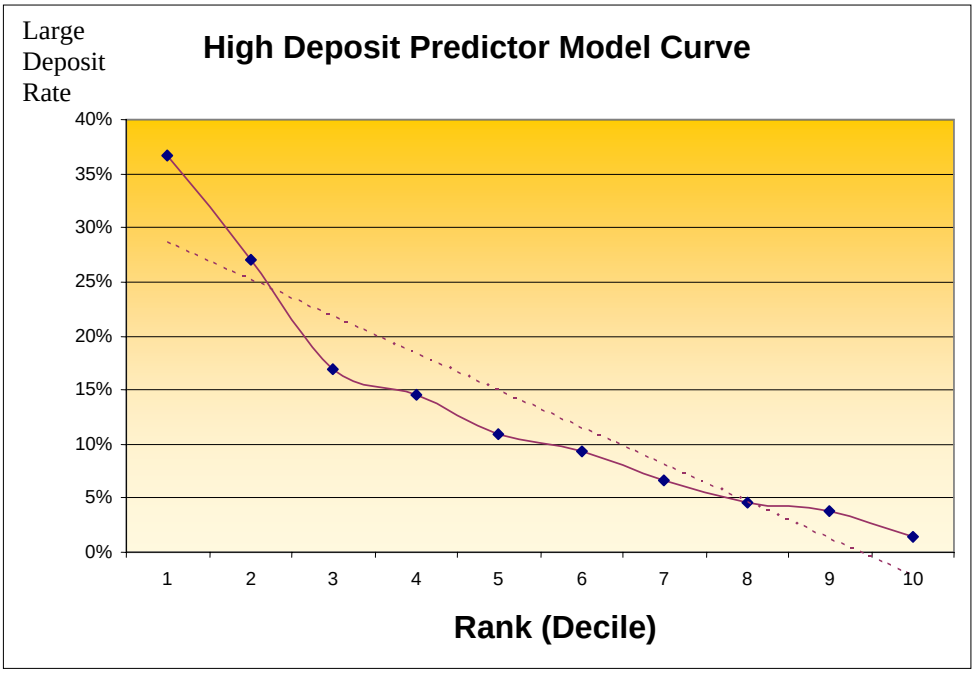
This particular variable indicates that as the % of a client's portfolio in registered assets relative to total assets becomes larger, it is less likely that the client will make a large deposit.

With the correlations and EDA reports being completed, multiple iterations of regression exercises were conducted to produce the high deposit model. Listed below is the actual final model report.

SAS Variable Name	Variable Description	% Contribution to Model Equation	+ or -
RATIO_a_reg	Total Registered assets / Total individual assets	48.4%	-
Tenure	# Years since joined the CMA	20.6%	+
NET_SALE_NOW_TC	NET_SALE_NOW > 5643 = 1; else = 0	13.5%	+
a_ip_tc	Total Investment Portfolio Assets, where a_ip > 75077 = 1 ; else = 0	7.5%	+
RESP	$ in all Reg. Education Savings Plan accounts	4.4%	-
KYC_LIQA1	KYC on Classic+: Estimated Liquid Assets (1:UNDER$50K 2:$50,001-$100K 3:$100,001-$250K 4:$250,001-$500K 5:0VER$500K)	2.7%	+
SCORE1_INV	Investment Product Breadth score	1.7%	-
RATIO_t_ip2_tc	Ratio_t_ip2 > 0.036839 = 1 ; else = 0	1.4%	+

From the above report, we observe that the strongest variable in the model is the ratio of registered assets as a % of total assets. As this ratio increases, the likelihood of having a large deposit decreases. Conversely, customers or clients with higher tenure are more likely to make large deposits. Together these variables account for almost 70% of the model's explanatory power.

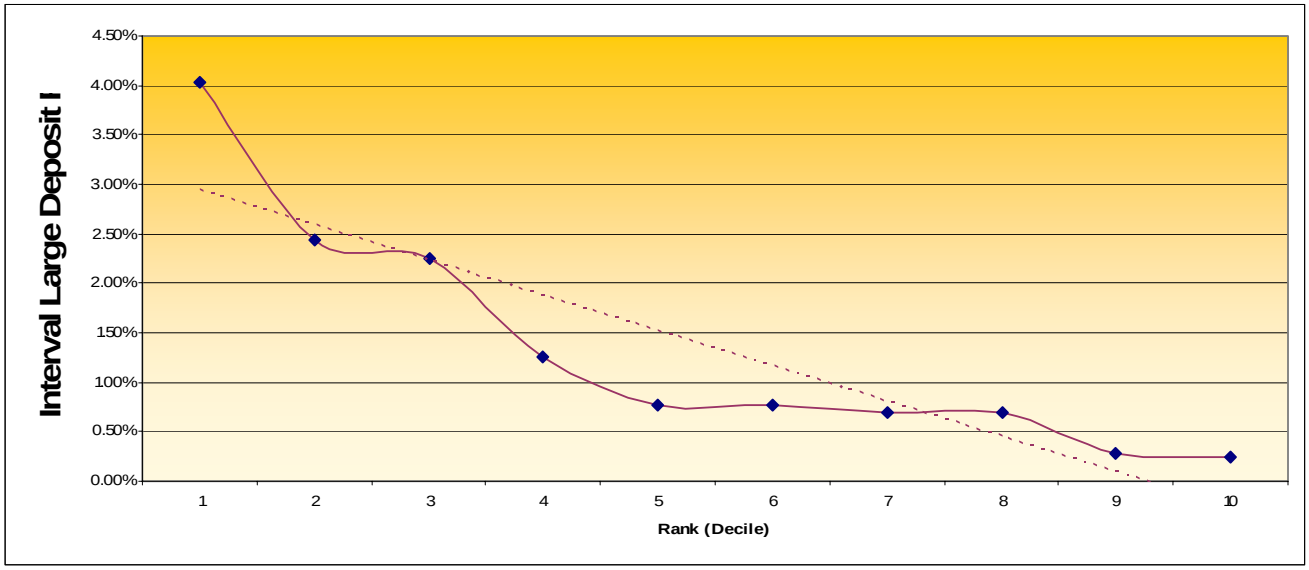
With the model developed, the next step was to validate our efforts on the holdout or validation sample.



Although these results demonstrated tremendous success when looking at the model's ability to rank order large deposits at a customer level, we were concerned with results being overstated. In fact, the lift ratio between the top decile and bottom decile is around 25 to 1(i.e. clients in the top decile were 25 times more likely to make a high deposit than clients in the lowest decile) - an extremely high lift number. Given the approach to developing the objective function and the actual model variables themselves however, we felt that this model would still perform very effectively, albeit perhaps not at the level as seen in the above Lorenz curve or trend line.

4. Actual In-Field results

When this model was applied in the field, the actual results were as follows:



In the above chart, the actual large deposit rate for the time period of analysis (3 months) was 1.34%. These results do indeed demonstrate the model's tremendous ability to rank order those investment clients most likely (decile 1) to make a high deposit to clients that are least likely (decile 10) to make a high deposit. The lift in application between these two deciles is close to 20 to 1. Generally, superior model performance occurs when a model can achieve a lift differential greater than 8 between the top decile and the bottom decile. A lift factor of 20 to 1 for this model certainly does demonstrate the model's superior performance.

5. Conclusion

Although our organization has been producing models for many years, the most challenging part of this exercise was iterative discovery process used to identify and define large deposits behaviour. The specific rules that we used to define large deposits will of course vary from organization to organization. But we wanted to demonstrate an approach or methodology in how we look at the data and information in arriving at our definitions. There are many situations involving modeling exercises where the objective function or behaviour to be predicted is not well defined. But a disciplined, iterative approach to looking at the data can provide the necessary insights to arrive at a relatively concise definition of the behaviour we are trying to predict.