WhitePaper

Multicollinearity in Customer Satisfaction Research

Jay L. Weiner, Ph.D.
Senior Vice President, Director of Marketing Sciences

Jane Tang
Vice President, Marketing Sciences

Ipsos Loyalty

About Ipsos Loyalty

Ipsos Loyalty is a global, specialized practice dedicated to helping companies improve business performance through customer satisfaction management, customer relationship management, and employee climate management. Ipsos Loyalty provides a state-of-the-art approach to customer-driven business performance through a modular suite of innovative research tools that provides an integrated framework to identify complex global business solutions. Ipsos Loyalty is an Ipsos company, a leading global survey-based market research group. To learn more, visit www.ipsosloyalty.com.

About the Contributors

Jay Weiner, Ph.D., Senior Vice President, Marketing Sciences

Jay consults with many leading corporations on marketing and market research issues. He specializes in applying advanced methods to help companies make better marketing and business decisions both domestically and globally. Jay has expertise in pricing, segmentation, customer and employee loyalty, conjoint analysis, and discrete choice analysis, in addition to solid multivariate statistical skills.

Jay has his doctorate in marketing research from the University of Texas at Arlington, as well as an MBA in finance and commercial banking and a BBA in marketing. Jay has published and presented numerous papers on conjoint, choice, and pricing research in conference proceedings.

Jane Tang, Vice President, Marketing Sciences

Jane provides expertise in analytical and methodological market research using various statistical techniques, from the basic univariate procedure to advanced multivariate. She is known for her research on analytical techniques and adaptation of techniques to the market research environment. Recently, her efforts have been concentrated in apex adjustment, Hierarchical Bayes models, discrete choice models, segmentation, and database marketing. She also serves as a sampling consultant for many project teams.

Jane has a B.Sc. and a M.Sc. in statistics from the University of Manitoba. She is also a graduate of the Sampling Program for Survey Statisticians from the Survey Research Center at the University of Michigan.

Abstract

This paper examines the strengths and weaknesses of four commonly used tools for modeling customer satisfaction data. Most customer satisfaction (CSAT) studies are plagued with multicollinearity, meaning that several of the independent causal variables are highly correlated, resulting in output that may cloak true drivers of satisfaction or dissatisfaction. When compounded by the fact that most CSAT studies are tracking studies, there is a significant challenge on how to model the data and deliver stable, actionable results to clients. As researchers and consultants, we must be sure that differences in results from one wave to the next are true differences in the market and not just, say, the result of a small number of respondents checking 8 instead of 7 on the last wave of a questionnaire.

The six traditional CSAT modeling techniques compared in this paper are:

1. Ordinary Least Squares
2. Shapley Value Regression
3. Penalty & Reward Analysis
4. Kruskal's Relative Importance
5. Partial Least Squares
6. Logistic Regression

The comparison begins with results that show the relative impact of multicollinearity on each technique, using a simulated data set. Then, results based on bootstrap samples pulled from this data set show the relative stability of the various techniques. Finally, a case study demonstrates how the various methods perform with a real data set.

Introduction

In customer satisfaction (CSAT) studies, we often conduct driver analysis to understand the impact of explanatory variables on the overall dependent variable. That is, we need to provide the client with a list of priority items that can be improved and that will have a positive impact on overall satisfaction or customer loyalty and retention. Typically, the goal is to establish the list of priorities and relative importance of the explanatory variables, rather than try to predict the mean value of customer satisfaction if these improvements were implemented. Since most CSAT studies are tracking studies, the results can be monitored over time to determine if the desired changes are occurring.

We must be sure that changes in the results are in fact customer response to the client's marketing efforts and not just phantoms of the analytic tool used to build the model. The latter often happens as a result of multicollinearity, which is a serious problem in many CSAT studies and presents two challenges in modeling CSAT data. The first is accurately reflecting the impact of several independent variables that are highly correlated. The second is insuring that the results are consistent wave to wave when tracking a market over time. This paper illustrates the problems that multicollinearity present in modeling data and then compares the results from the four aforementioned modeling techniques.

The Issue of Multicollinearity

In market research, multicollinearity can be controlled or altogether avoided by a well-designed questionnaire. For most researchers, this is a common desire, but difficult to achieve. In most CSAT studies, we measure a variety of attributes that are often highly correlated with each other. For example, in evaluating the service provided by a customer call center, we frequently ask respondents to rate satisfaction with the friendliness of the operator, and also of the operator's ability to handle the problem the first time. We often see that these two attributes are highly correlated with each other. This may be due to halo effects in that most customers that are happy with the resolution of the problem will reflect back and state that the operator was friendly. Regardless of the reason for the correlation between these two attributes, we need to find a modeling tool that is capable of determining the relative contribution of each of these attributes to overall satisfaction.

To set up the comparison, we created a data set with 5,000 observations that is typical of CSAT studies, where the properties of the dependent and independent measures are known. The simulated data set has two pairs of independent variables and a dependent measure (overall satisfaction). The attributes in the first pair (q^1 and rq^1) of independent measures are constructed to be almost perfectly correlated with each other and highly correlated with the dependent variable. The attributes in the second pair (q^2 and rq^2) are also highly correlated with each other, but less correlated with the dependent measure. All variables are on a 10-point rating scale. The correlation matrix is shown in Figure 1.

Figure 1: Correlation Matrix for Wave One

OS = Overall Satisfaction, the dependent measure

	OS	q^1	rq^1	q^2	rq^2
OS	1.00	0.63	0.62	0.39	0.38
q^1	0.63	1.00	0.98	0.26	0.26
rq^1	0.62	0.98	1.00	0.25	0.25
q^2	0.39	0.26	0.25	1.00	0.98
rq^2	0.38	0.26	0.25	0.98	1.00

Impact of Multicollinearity on Ordinary Least Squares Regression

A common modeling tool is ordinary least squares regression (OLS).
If we regress Overall Satisfaction on q^1 and q^2, we find the following results.

Figure 2: Ordinary Least Square Regression Output

Variable	Beta Coefficient	P-value
q^1	0.57	<.0001
q^2	0.23	<.0001

Both variables are significant, as indicated by the P-values, and have a positive impact on overall satisfaction, as indicated by the correlations, or beta coefficients. These results are consistent with the correlation matrix shown in Figure 1.
To demonstrate the impact of multicollinearity on OLS, we ran the model adding rq^1 and rq^2.

Figure 3: Impact of Multicollinearity on OLS when Variables Known to be Highly Correlated Are Added

Variable	Beta Coefficient	P-value
q^1	0.51	<.0001
rq^1	0.06	0.19
q^2	0.23	<.0001
rq^2	0.00	0.97

Because q^1 and rq^1 are almost perfectly correlated, the regression coefficients in Figure 3 show that q^1 has more than eight times the impact on overall satisfaction than rq^1 (.51 vs. .06) while the other is not significant, as indicated by the high P-value. If the model were run using only rq^1 and rq^2, we would see results very similar to the model with q^1 and q^2 in Figure 2.

The second challenge multicollinearity presents in CSAT research is in analyzing the results in subsequent periods of time. As mentioned, most CSAT studies are tracking studies. We execute the first wave to understand where we are, then we track our key measures over time to measure the client's effectiveness in improving overall satisfaction. To simulate the impact of multicollinearity on the ability to model subsequent waves of data, we created a second data set to represent a second wave of data collection using the same variables (Figure 4).

Figure 4: Correlation Matrix Wave Two

OS = Overall Satisfaction

	OS	q1	rq1	q2	rq2
OS	1.00	0.64	0.63	0.37	0.38
q1	0.64	1.00	0.98	0.25	0.25
rq1	0.63	0.98	1.00	0.24	0.24
q2	0.37	0.25	0.24	1.00	0.98
rq2	0.38	0.25	0.24	0.98	1.00

In this wave of data, rq^2 is slightly more correlated with overall satisfaction than q^2. When we run the OLS, we see the following results.

Figure 5: Impact of Multicollinearity: Same Variables, Different Point in Time

Variable	Beta Coefficient	P-value
q1	0.58	<.0001
rq1	0.01	0.78
q2	0.01	0.89
rq2	0.21	<.0001

If these were the results of an actual tracking study, we would have advised the client to focus on the attribute q^2 in wave 1, but in wave 2, we would have encouraged the client to focus on the attribute rq^2. It would be nice to know that the change of focus is due to actual events that have happened in the market and not due to a small change in the overall correlation matrix. Clearly there are risks when using OLS in CSAT research.

Other Tools in the Toolbox

There are other models that consider all possible combinations of explanatory variables. The conclusions we draw from these models refer to the usefulness of including any attribute in the model and not its specific impact on improving attributes like overall satisfaction. These models include: Kruskal's Relative Importance (KRI), Shapley Value (SV) Regression, and Penalty & Reward Analysis. Unlike traditional regression tools, these techniques are not used for forecasting. For example, in OLS, we predict the change in overall satisfaction for any given change in the independent variables. The tools addressed in this section are used to determine how much better the model is if we include any specific independent variable versus models that do not include that measure.

For each of these modeling techniques, the impact of a particular variable on the dependent variable is determined by its contribution to the model, measured by the difference in XXX between including it in the model and not including in the model. The XXX varies by technique.

Kruskal's Relative Importance

For Kruskal's Relative Importance, OLS regression is used for all possible combinations of explanatory variables. Here, the contribution of each attribute, or XXX, is measured by squared partial correlations. KRI can be run using SAS with PROC REG with selection=rsquare & PROC IML.

$$KRI_j = \sum_k \sum_i \frac{k!(n-k-1)!}{n!} \left[(SSE_{i|j} - SSE_{i|j(-j)}) / SSE_{i|j} \right]$$

where $SSE_{i|j}$ is the SSE of a model i containing predictor j

$SSE_{i|j(-j)}$ is the R^2 of a model i without j

Including q^1 and q^2 in the model, we see that the results are very consistent with the results obtained using OLS in that q^1 has about three times the impact of q^2.

Figure 6

	Wave 1	Wave 2
q^1	0.37	0.39
q^2	0.12	0.11

When we add the highly correlated independent measures to the model, Kruskal's Relative Importance essentially splits the importance for each pair of independent measures. Both q^1 and rq^1 are equally important and more than three times more important than q^2 and rq^2. This result often has more face validity with clients than the results obtained using OLS because the correlations present empirical evidence, and it overcomes concerns about certain variables appearing insignificant in the model.

	Wave 1	Wave 2
q^1	0.20	0.21
rq^1	0.18	0.19
q^2	0.06	0.06
rq^2	0.06	0.06

Shapley Value Regression

Like Kruskal's Relative Importance, Shapley Value regression uses OLS regression for all possible combinations of explanatory variables. In Shapley Value regression, the contribution of each attribute, or XXX, is measured by the improvement in R-square. This can be executed with SAS using PROC REG with selection=rsquare & PROC IML.

$$SV_j = \sum_k \sum_i \frac{k!(n-k-1)!}{n!} [\upsilon(M_{i|j}) - \upsilon(M_{i|j(-j)})]$$

where $\upsilon\ (M_{i|j})$ is the R^2 of a model i containing predictor j
$\upsilon\ (M_{i|j(-j)})$ is the R^2 of a model i without j

Again, a model with just q^1 and q^2 is similar to the results obtained using OLS. The results when we add the highly correlated variables to the model are almost identical to those seen with Kruskal's Relative Importance.

Figure 7

	Wave 1	Wave 2
q^1	0.18	0.19
rq^1	0.17	0.17
q^2	0.05	0.05
rq^2	0.05	0.05

	Wave 1	Wave 2
q^1	0.35	0.36
q^2	0.10	0.10

Penalty & Reward Analysis

Penalty & Reward Analysis is unique in that instead of a single linear model, we build two models. The overall sample is divided into the delighted group (e.g. rated 10, 9 or 8 in overall satisfaction on a 10-point scale) and the less than satisfied group (e.g. rated 3, 2, or 1). Dissatisfaction with explanatory variables is a potential source of penalty or barrier to satisfaction, while delight with explanatory variables is a potential source of reward or driver of delight. This approach is useful when there is a suitable normal distribution of the overall dependent variable and suspect the drivers of satisfaction are likely different from the drivers of dissatisfaction. The unique contribution of each attribute is measured by the difference in overall total unduplicated reach by each combination of the explanatory variables between the delighted group and the less than satisfied group. The importance of a variable is measured as the sum of its penalty and reward. This analysis is done using SAS IML.

The results for the Penalty & Reward Analysis are very similar to those obtained with Kruskal's Relative Importance and Shapley Value regression.

Figure 8

	Wave 1	Wave 2
q^1	0.70	0.70
q^2	0.44	0.43

	Wave 1	Wave 2
q^1	0.36	0.36
rq^1	0.36	0.35
q^2	0.23	0.22
rq^2	0.22	0.21

Case Example

The dataset is from an Ipsos Loyalty Optimizer™ study. The dependent measure for the models presented is the Security Index, which is a composite measure of customer loyalty and is a continuous variable. For some models (such as for Penalty & Reward Analysis and logistic regression), the data can be divided into the high vs. low loyalty groups. Each respondent only evaluates one company (single observation). There are five constructs each with 2 attributes: Relationship, (feel like a friend, trust and treat fairly), Experience (quality, satisfaction), Offer (relevance/differentiation), Brand (familiarity/popularity), Price (willingness to pay/comparison to competitors). All explanatory variables are expected to have a positive impact on the dependent variable.

Objective

The goal is to assess the stability of the driver analysis from repeated samples of different sample sizes and to compare results using OLS, Kruskal's Relative Importance, Shapley Value regression and Penalty & Reward Analysis.

Methodology

We created several datasets by pulling repeated bootstrap samples from the original Loyalty Optimizer data, e.g., 100 bootstrap samples (i.e. sampling with replacement) for each sample sizes of: n=2,400 (original sample size), n=1,000, n=500, n=300 and n=100.

Using the Security Index variable as the dependent variable we ran OLS regression, STEPWISE OLS, Partial Least Square Regression (PLS), Kruskal's Relative Importance, Shapley Value regression, Penalty & Reward Analysis, and compared the results.

For OLS, all parameters are entered into the model. In the stepwise method of OLS, parameters that are not significant in the model are assumed to be zero. For logistic regression and Penalty & Reward Analysis we compared results using the High vs. Low loyalty category as the dependent measure.

Criterion for Comparison

First, we examined the gaps in the importance calculation for each attribute between the first two bootstrap samples in each sample size. Note, as all attributes are assumed to have a positive impact, negative betas are set to "0." The resulting betas are re-proportioned, such that the sum of importance equals 100%. The following table shows the results of the OLS run on the full sample size of 2,400 respondents. The average gap in importance between wave 1 and wave 2 is 4.0%, the largest difference is at 8.7%

Figure 9

The beta column is the computed coefficient. If the coefficient is negative, we set it to zero. These are highlighted. The importance is computed from the Beta value, and the gap represents the difference between importance scores between the two samples. In a perfect world, the gaps would all be zero.

n=2,400	Sample 1			Sample 2			
OLS Regression	Beta	"zeroed"	Importance	Beta	"zeroed"	Importance	Gap
Feels like a friend	0.00	0.00	0%	0.00	0.00	2%	2%
Trust and treat fairly	0.01	0.01	10%	0.01	0.01	6%	4%
Quality	0.00	0.00	0%	0.01	0.01	5%	5%
Satisfaction	0.04	0.04	34%	0.03	0.03	31%	4%
Relevance	0.03	0.03	33%	0.03	0.03	24%	9%
Differentiation	0.00	0.00	3%	0.00	0.00	1%	1%
Familiarity	0.01	0.01	5%	0.01	0.01	13%	8%
Popularity	0.00	0.00	0%	–0.01	0.00	0%	0%
Willing to pay	0.01	0.01	12%	0.01	0.01	9%	3%
Comparison to competitor's price	0.00	0.00	3%	0.01	0.01	7%	4%
Average Gap between Sample 1 & 2							4.0%
Maximum Gap between Sample 1 & 2							8.7%

The second step in the comparison is to calculate the Coefficient of Variation by dividing the standard deviation by the mean across 100 bootstrap samples. The techniques with lower gaps and median coefficients of variation percentages are more stable, as shown in Figure 11.

Figure 10

Mean = average coefficient across the 100 samples

n=2,407 OLS Regression	Mean	STD DEV	C of V
Feels like a friend	0.004	0.004	1.094
Trust and treat fairly	0.004	0.004	1.068
Quality	–0.001	0.005	4.127
Satisfaction	0.038	0.005	0.136
Relevance	0.031	0.004	0.140
Differentiation	0.005	0.003	0.699
Familiarity	0.015	0.004	0.275
Popularity	–0.007	0.003	0.484
Willingness to pay	0.015	0.002	0.148
Comparison to competitor's price	0.003	0.003	1.278
		Median C of V.	0.592

Figure 11: Summary of all methods.

	Dependent variable: Security Index					Dependent variable: High vs. Low loyalty		
Sample size	OLS	OLS (stepwise)	PLS	KRI	SV	Logistic Reg	Logistic Reg (stepwise)	P&R
Average Gap Between Sample 1 & 2								
2,400	4.0%	5.5%	3.8%	0.7%	0.6%	3.9%	6.1%	0.6%
1,000	2.7%	2.2%	2.4%	1.5%	1.4%	3.0%	4.2%	0.8%
500	5.6%	6.4%	5.3%	3.2%	2.8%	4.9%	11.5%	1.6%
300	6.7%	8.0%	7.6%	3.8%	2.9%	5.4%	6.7%	2.3%
100	9.9%	20.0%	9.5%	3.6%	3.4%	–	–	–
Maximum Gap Between Sample 1 & 2								
2,400	8.7%	13.8%	8.0%	2.3%	2.1%	10.2%	16.4%	1.5%
1,000	6.4%	7.7%	5.6%	3.5%	3.3%	7.1%	13.5%	1.3%
500	13.6%	13.6%	12.2%	6.8%	5.3%	13.1%	26.8%	3.4%
300	30.5%	33.8%	32.5%	11.1%	7.3%	11.3%	33.5%	6.7%
100	21.1%	44.9%	18.7%	10.2%	9.6%	-	-	-
Median Co-efficient of Variation								
2,400	59%	97%	57%	13%	12%	73%	193%	7%
1,000	75%	115%	73%	19%	17%	60%	143%	11%
500	67%	97%	64%	22%	20%	110%	236%	13%
300	112%	140%	111%	32%	33%	184%	341%	17%
100	172%	168%	166%	38%	39%	–	–	–

Conclusions

With large sample sizes, all the methods perform fairly well. If the objective is to establish relative importance, rather than forecasting, methods that take into consideration all the possible combinations of the explanatory variables (i.e. Kruskal's Relative Importance, Shapley Value Regression, and Penalty & Reward Analysis) are worthwhile since they have much smaller sample-to-sample variability. These methods also do not collapse as quickly as the sample size decreases. This is more suitable where we don't have large sample.

Tools using all possible combinations of attributes do a better job of dealing with issues of multicollinearity. OLS tends to understate the impact of highly correlated attributes and can in fact suggest a negative coefficient in some models for attributes that are expected to have a positive impact on the dependent measure. Conversely, Kruskal's Relative Importance, Shapley Value Regression, and Penalty & Reward Analysis assign equal weights to attributes that are highly correlated with each other. When applied to tracking data, these tools are very stable in predicting the impact of attributes between waves. We have more confidence that changes in the impact of independent measures are in fact true differences in the market and not due to small changes in the correlations with the dependent measure.

References:

Lipovetsky, S., and Conklin, M. (2001) "Analysis of Regression in Game Theory Approach," *Applied Stochastic Models in Business and Industry,* 17, 319 – 330

Kruskal, W. (1987). "Relative Importance by averaging over orderings," *The American Statistician,* 41(1)

Related Readings

To build your knowledge of the technical and strategic aspects of Loyalty research visit our Knowledge Centre at www.ipsosloyalty.com/knowledgecenter.

North America

Europe

Latin America

Asia-Pacific

Middle East