An Analysis of Customer Preference of Automobile Products using SAS®

Akkarapol Sa-ngasoongsong and Satish T.S. Bukkapatnam
Oklahoma State University, Stillwater OK, 74078, USA

ABSTRACT

This paper presents an analysis of customer preference of automobile products in terms of Willingness-To-Pay (WTP) of vehicle attributes using SAS®, including customer segmentation, and dimensionality reduction of WTP vector. This analysis addresses issues related to the preparation of customer inputs for effective prediction of customer preferences. Our analysis shows that respondents show maximum variability in specifying their WTP for Body Type (BT) attribute and hence classification based on BT associations is the most suitable way to differentiate between respondents from available data. In case of dimensionality reduction of WTP vector, we propose a twofold criteria for WTP rankings. The first addresses the importance of WTP for an attribute level from a respondent's utility maximization standpoint. The second criteria addresses the predictability of WTP from the available data using Signal-to-Noise Ratio (SNR). The combination of both criteria provides excellent insights to the analyst regarding importance and predictability of WTP data, and also provides a tractable way to the dynamic modeling of customer preferences.

INTRODUCTION

Early stages of product development hinge heavily on understanding the needs of the potential customers. This helps the firm in optimally positioning a single new product or a line of new products constituting a portfolio to perform well on metrics such as contribution margin or revenue in a competitive market space. Krishnan and Ulrich [1] present an excellent review of literature on product development practice and reflect that frameworks and decisions adopted by the firms vary with the product offering. This is indeed true in cases of fast moving consumer goods. The entire exercise of generating a new concept can be achieved in a small horizon. Many of existing techniques [2, 3] that have been used in choice modeling approach for understanding customer needs assume time invariant behavior of the customer. These methods seldom consider the evolving nature of customer choice over time which can have a significant impact on market performance.

For product offerings where the product development window is quite large as is the case of an automobile where the concept-to-release time takes approximately 36-48 months, one would expect the customer preferences to have evolved during this period. As a result the vehicle portfolio designed based on past customer preferences would be suboptimal relative to market shares in the future and hence the need to capture the dynamics of the customer preferences arises. Oftentimes, customers preferences are expressed in terms of willingness-to-pay (WTP) [4] metric, which refers to the maximum monetary amount a consumer is willing to pay for a good or service. In this paper, we addresses issues related to the preparation of customer inputs for effective prediction of customer preferences. The specific issues addressing includes customer segmentation and dimensionality reduction of the WTP vector.

The WTP data in this paper has been obtained from automobile customers using self-explicated survey technique. Information was collected from customers on 25 vehicle attributes and accounting for all possible levels of the attributes. The size of the WTP vector including attribute levels for each individual customer is 144. The details of each vehicle attribute and attribute levels can be found in Appendix A.

CUSTOMER SEGMENTATION AND DIMENSIONALITY REDUCTION

The Block diagram summarizing the analysis is presented in Figure 1. As mentioned in the introduction section, the analysis results addressed in this paper will be used to prepare the customer inputs for effective prediction of customer preferences. Given the confounded nature of the WTP data, one would intuitively expect to observe the presence of multicolinearity amongst the various attribute dimensions for each respondent. As an example, respondents opting for and hence revealing higher WTP's for a particular Body Type & Size (BTS) level would correspondingly have shown higher WTP's for certain levels of roominess attributes. Therefore, the multicolinearity reduction technique is used to reduce multicolinearity and identify attribute dimensions showing maximum variations. Customer segmentation is based on this set of attribute dimensions. Further, having clustered respondents, WTP for attribute levels which show significance and variation over time is identified for each respondent cluster and will be used for customer preferences prediction.

Multicolinearity Reduction

The objective in this analysis is to condense information contained in a large number of WTP variables into a smaller set of variables to be used in subsequent analysis (Prediction, Clustering, etc.) Common techniques to eliminate some redundant variables are, for example, Principal Component Analysis (PCA) and Exploratory Factor Analysis (EFA). PCA creates the weighted linear combinations of WTP variables while retaining most of the variability in the data. The advantage of this method is the fewer variables and no multicollinearity among the variables. In contrast with PCA, EFA assumes that observed WTP variables are linear combinations of (or, caused by) underlying common factors and specific factors (errors). The results of this analysis can be used for clustering customers.

Figure 1. Process Diagram of Analysis

To present the customer segmentation for WTP data using PCA and EFA, the WTP data is used to create the principal components in SAS® Enterprise Guide® 4.3. Table 1 shows the Eigenvalues of the correlation matrix of 144 WTPs. The results of Eigenvalues of the correlation matrix of 144 WTPs show that the first Eigenvalue is 37.367 which indicates the variance explained by the first principal component. The percentage of variance explained by first Eigen value (first PC) is 25.95%. If we use rule of thumb (selecting PC's with Eigenvalue >1), only 20 PCs will be selected. These uncorrelated 20 PCs can capture 80.85% of the total variance in this data set. Another way to find the number of PCs to retain is using the scree plot as shown in Figure 2. Since we have a large number of original variables (144), it is very difficult to identify the elbow of this scree plot. From Figure 2, the elbow of the scree should be between 15 to 25 PCs.

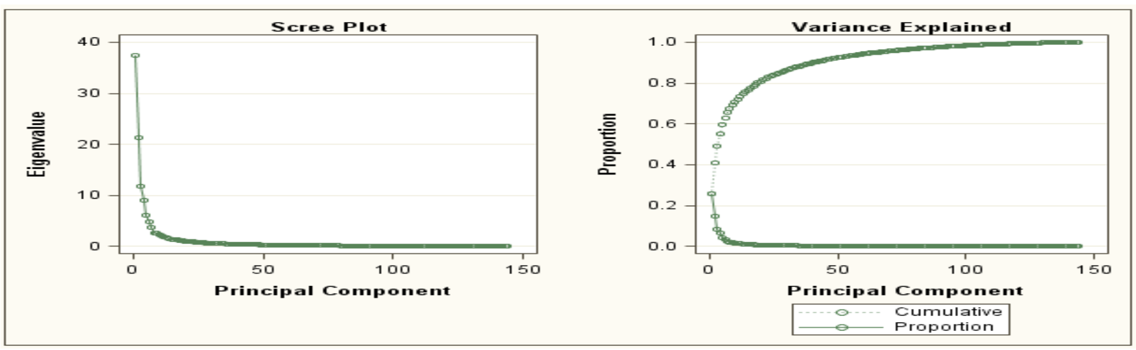


Figure 2. Scree plot and Variance Explained by Principal Components

Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	37.3666	16.0566	0.2595	0.2595
2	21.3100	9.4683	0.148	0.4075
3	11.8416	2.7251	0.0822	0.4897
4	9.1165	3.0378	0.0633	0.553
5	6.0787	1.1701	0.0422	0.5952
6	4.9086	1.1830	0.0341	0.6293
7	3.7256	1.0369	0.0259	0.6552
8	2.6887	0.0643	0.0187	0.6739
9	2.6244	0.3743	0.0182	0.6921
10	2.2502	0.2347	0.0156	0.7077
11	2.0155	0.1511	0.014	0.7217
12	1.8644	0.1223	0.0129	0.7347
13	1.7421	0.1712	0.0121	0.7468
14	1.5709	0.1440	0.0109	0.7577
15	1.4269	0.1029	0.0099	0.7676
16	1.3241	0.0717	0.0092	0.7768
17	1.2523	0.0768	0.0087	0.7855
18	1.1756	0.0808	0.0082	0.7936
19	1.0948	0.0484	0.0076	0.8012
20	1.0464	0.0494	0.0073	0.8085
21	0.9970	0.0326	0.0069	0.8154
22	0.9644	0.0392	0.0067	0.8221
23	0.9251	0.0521	0.0064	0.8285
24	0.8730	0.0782	0.0061	0.8346
25	0.7948	0.0482	0.0055	0.8401
26	0.7466	0.0364	0.0052	0.8453
27	0.7102	0.0466	0.0049	0.8502
28	0.6636	0.0061	0.0046	0.8549
29	0.6575	0.0393	0.0046	0.8594
30	0.6183	0.0237	0.0043	0.8637

Table 1. Eigenvalues of the Correlation Matrix

Figure 3 shows the component scores matrix of the principal components. The histograms in the diagonal line show the distribution of the first 10 PCs. The scatter plots in the matrix show the plots between one PC to the others. As expected, there is no pattern in all scatter plots in the component scores matrix (because these PCs are uncorrelated), and no consistent set of outlier seem to be evident.

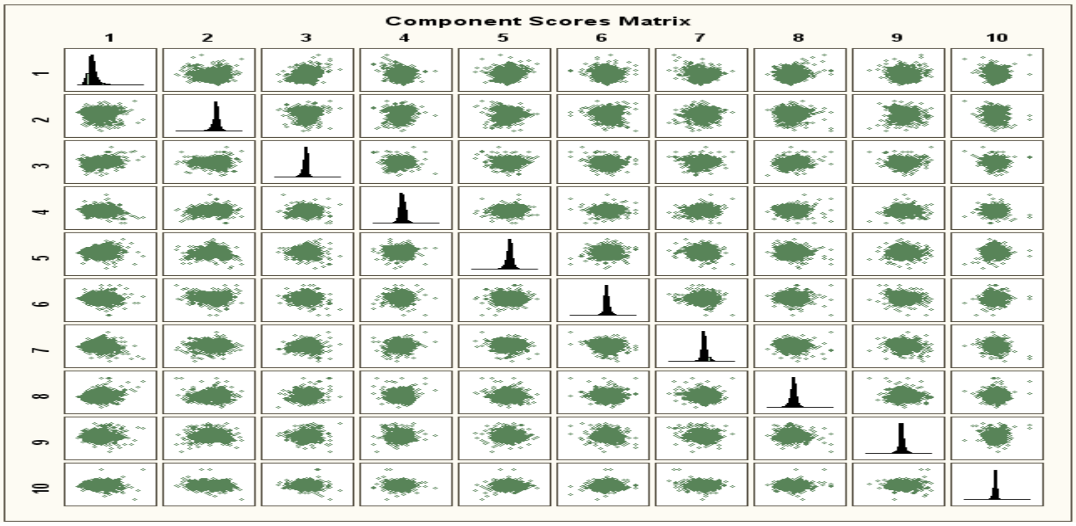


Figure 3. Component Scores Matrix of the First 10 PCs

We conclude that PCA reveals that the 144 respondent dimension data can be best explained using the 20 PC's obtained as this reduced set explains 80% of the variation observed across respondents. As these PC's are uncorrelated one could perform respondent clustering in the reduced 20 dimensional space. However doing so would require us to operate entirely in the PC space which is not desired as some of the important attributes which may be uncorrelated with other attributes may get eliminated.

Clustering is required as an intermediate step in our analysis, using which we retain customer heterogeneity and observe aggregate WTP's at cluster level over time. Given this we next try to identify attributes which exhibited maximum variability across respondents. This attribute or set of attributes would be further used for clustering customers using EFA. In order to identify this attribute or set of attributes we need to analyze the salient WTP attributes in each PC. We report the score or coefficients ranking of WTP for attribute levels for the first three PCs in Table 2. Interestingly, the top 10 important WTPs in each of the first three PCs have some meaningful patterns which are different from PC to PC.

Rank	PC1	Score	PC2	Score	PC3	Score
1	WTP_73	0.1405	WTP_48	0.1888	WTP_17	0.1952
2	WTP_67	0.1405	WTP_60	0.1881	WTP_15	0.1847
3	WTP_72	0.1387	WTP_44	0.1866	WTP_25	0.1842
4	WTP_66	0.1356	WTP_46	0.1843	WTP_9	0.1770
5	WTP_115	0.1345	WTP_47	0.1836	WTP_10	0.1738
6	WTP_68	0.1343	WTP_43	0.1831	WTP_28	0.1733
7	WTP_75	0.1329	WTP_45	0.1822	WTP_13	0.1680
8	WTP_70	0.1325	WTP_59	0.1799	WTP_11	0.1658
9	WTP_124	0.1324	WTP_58	0.1769	WTP_32	0.1646
10	WTP_118	0.1323	WTP_54	0.1696	WTP_23	0.1644

Brand: WTP_6 to WTP_41, BTS: WTP_42 to WTP_64, Comfort & Convenience: WTP_65 to WTP 76, Safety Features: WTP_114 to WTP_125, In Appendix A we provide a map of AMIS WTP's to corresponding attribute levels.

Table 2. WTP Ranks on Loadings of the First three PCs

For the first PC, seven WTPs (from top 10) are WTPs for comfort & convenience attribute. The other three are WTPs for safety features attribute. This indicates that comfort & convenience, and safety features attributes are the most significant attributes for PC1. All of the top 10 WTPs for second PC are WTPs for Body Type & Size (BTS). Therefore, BTS is the most significant attribute for PC2. For PC3, all of the top 10 WTPs are WTPs for Brand.

As WTP values for each respondent define a dollar amount the respondent is willing to pay for a particular level, one would intuitively expect feature based levels to show maximum variability amongst respondents as it directly reflects price. This explains the presence of convenience & comfort features and safety based features in PC1. By leaving out PC1 we eliminate effects due to presence of price information in WTP and only consider attributes which inherently show variation in tastes across respondents devoid of what they are priced at.

The results of principal component score or coefficient thus show that the most significant attributes in terms of capturing the variation in the data are BTS and Brand. However, as we discussed this earlier in this section as clustering at a granular level is not the only goal of our analysis we only retain the BTS attribute from PCA. The next step is to use factor analysis to identify latent variables if any which underlie the process or behavior of the respondents and explain their associations to BTS groups. Doing this would reduce the number of BTS groups further, thereby resulting in smaller cluster number.

Customer Segment Classification

Table 3 shows the preliminary Eigenvalues from factor analysis of 23 WTPs for the BTS attributes of all respondents considered in PCA. From the Preliminary Eigenvalues table, the values of first 13 factors are positive. Therefore, the first 13 factors are retained by default. The criterion specifies the smallest Eigenvalue for retaining a factor. The traditional rule of Eigenvalue >1 is less meaningful in determining the number of factors to retain because we have negative Eigenvalues for factor analysis

A better way to find the number of factors to retain, and to understand the structure of the data is to look at the Rotated Factor Pattern of the 13 factors as shown in Table 4. Rotated factors demonstrate the presence of simple structure in the data. The interpretation and naming of the factors are based on factor loadings. From the Rotated Factor Pattern, we generate a table of dominant WTP's for each factor in Table 5.

Preliminary Eigenvalues: Total = 160.970627				
Average = 6.99872292				
	Eigenvalue	Difference	Proportion	Cumulative
1	139.0597	131.1523	0.8639	0.8639
2	7.9074	2.1991	0.0491	0.913
3	5.7083	2.4555	0.0355	0.9485
4	3.2528	0.5688	0.0202	0.9687
5	2.6839	0.7884	0.0167	0.9853
6	1.8955	0.3353	0.0118	0.9971
7	1.5602	0.5935	0.0097	1.0068
8	0.9667	0.2997	0.006	1.0128
9	0.6671	0.2359	0.0041	1.017
10	0.4311	0.1427	0.0027	1.0196
11	0.2885	0.0837	0.0018	1.0214
12	0.2047	0.1552	0.0013	1.0227
13	0.0495	0.1152	0.0003	1.023
14	-0.0656	0.1350	-0.0004	1.0226
15	-0.2006	0.0334	-0.0012	1.0214
16	-0.2339	0.0580	-0.0015	1.0199
17	-0.2919	0.0842	-0.0018	1.0181
18	-0.3761	0.0401	-0.0023	1.0158
19	-0.4162	0.0082	-0.0026	1.0132
20	-0.4244	0.0611	-0.0026	1.0105
21	-0.4856	0.0942	-0.003	1.0075
22	-0.5797	0.0509	-0.0036	1.0039
23	-0.6307		-0.0039	1

Table 3. Preliminary Eigenvalues for Factor Analysis

As our central goal in using EFA is to study the underlying structure of the BTS WTP's i.e. identify any underlying reduced set of explanatory latent variables, we group the BTS variables using factor loading shown in Table 5. A casual glance at Table 5 again reveals classification resulting due to price, for example, factors 4 & 8 comprise Sedan as a whole and factors 5 & 7 would represent the SUV segment. If we were to neglect the effects of price, these factors could be clubbed together and the 23 BTS variables could be explained using 8 latent variables. We conclude this subsection noting that respondents would be grouped by their Body Type (BT) preferences and the evolution of time aggregates at a BT level would be modeled for with respect to the evolving exogenous space.

To validate our conclusions regarding classifying respondents by BT associations we next perform a Canonical Discriminant Analysis (CDA) to identify significance level of WTPs that are used to discriminate groups of customers. Based on our forgoing conclusions, one would intuitively expect the WTP for BTS attribute levels to have a major role in discriminating amongst respondents. Unlike PCA, CDA analysis identifies the directions along which class-wise differences (separations) are largest in the attribute space and the effects of price observed when performing PC and EFA should not occur when performing the CDA. In this part, the analysis results have been created using WTP data for the sedan segment.

Rotated Factor Pattern													
	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7	Factor8	Factor9	Factor10	Factor11	Factor12	Factor13
WTP_MEAN_43	0.88059	0.28104	0.18370	0.12578	0.07557	0.12255	0.13074	0.06986	0.02283	-0.04749	-0.16602	0.01648	-0.15515
WTP_MEAN_44	0.86706	0.25993	0.13844	0.12040	0.17436	0.14501	0.10270	0.09300	0.01600	-0.01605	0.02917	0.00867	0.05420
WTP_MEAN_48	0.85442	0.32086	0.15701	0.15491	0.14348	0.14086	0.14849	0.11107	0.01318	0.04549	-0.08383	0.00069	0.19884
WTP_MEAN_46	0.84091	0.21816	0.14490	0.09671	0.23666	0.12661	0.06007	0.10993	0.00778	-0.02040	0.34146	0.00876	-0.01499
WTP_MEAN_45	0.83161	0.23842	0.21984	0.08790	0.15727	0.12856	0.07286	0.09807	0.02337	-0.04371	0.11270	-0.01884	-0.11683
WTP_MEAN_47	0.81931	0.34332	0.16068	0.15737	0.09058	0.13441	0.19387	0.07949	0.01488	0.03510	-0.12564	-0.01035	0.06816
WTP_MEAN_60	0.75103	0.36384	0.12575	0.15200	0.27172	0.11598	0.16358	0.10635	0.02516	0.35933	-0.00956	0.01022	0.01801
WTP_MEAN_58	0.64845	0.24111	0.13580	0.09139	0.53909	0.16379	0.11095	0.09404	-0.00168	0.12727	0.05993	0.07899	0.00825
WTP_MEAN_54	0.60422	0.31691	0.25248	0.14225	0.12751	0.16991	0.50944	0.07708	0.04722	0.03654	0.00242	-0.06769	0.00544
WTP_MEAN_59	0.55510	0.43356	0.20312	0.03500	0.33174	0.05040	0.11430	0.12419	0.15184	0.22779	-0.02386	-0.01491	-0.00474
WTP_MEAN_61	0.34634	0.80263	0.16385	0.17206	0.08403	0.04478	0.13455	0.07106	0.05868	0.00677	0.00850	0.09628	-0.00042
WTP_MEAN_64	0.33509	0.76059	0.20345	0.19215	0.13780	0.11705	0.09115	0.17801	-0.01770	0.02319	0.00672	-0.10106	-0.00616
WTP_MEAN_63	0.35535	0.73043	0.06915	0.34401	0.04589	-0.07229	0.16217	-0.06624	-0.18563	0.00083	0.01603	-0.02252	0.01180
WTP_MEAN_62	0.38995	0.68675	0.21748	0.03032	0.25466	0.10853	0.05373	0.23341	0.37467	0.05112	0.00813	-0.00532	0.00152
WTP_MEAN_56	0.39902	0.25002	0.73075	0.01767	0.35825	0.13334	0.07773	0.12298	0.06619	0.01418	0.00630	-0.11251	0.00431
WTP_MEAN_55	0.35782	0.42115	0.62755	0.14055	0.07103	0.09924	0.25511	0.04241	-0.03292	0.03130	0.01225	0.15609	-0.00183
WTP_MEAN_49	0.15837	0.23670	0.02477	0.78299	0.02881	-0.01910	0.06826	0.05311	0.00917	0.00828	0.00417	0.00824	0.00582
WTP_MEAN_50	0.16333	0.19959	0.13807	0.64055	0.04123	0.35723	0.05489	0.56342	-0.00805	0.02483	-0.00119	-0.02705	-0.01676
WTP_MEAN_57	0.51730	0.18252	0.31002	0.02583	0.72379	0.14270	0.04997	0.11709	0.02209	-0.02772	-0.00086	-0.02362	-0.00124
WTP_MEAN_42	0.08747	0.04393	0.01307	0.23220	0.01485	0.64346	0.11787	0.07158	0.00451	-0.00568	-0.01265	-0.11470	0.00185
WTP_MEAN_52	0.13603	0.00451	0.07631	-0.13466	0.08052	0.56134	-0.04241	0.05832	0.00814	0.01601	0.01562	0.09454	-0.00067
WTP_MEAN_53	0.54944	0.36015	0.21138	0.14831	0.07355	0.06342	0.65312	0.04718	-0.00879	0.02144	0.00547	0.02700	-0.00034
WTP_MEAN_51	0.34722	0.22873	0.11519	0.25637	0.26223	0.32783	0.05860	0.51606	0.06219	0.01879	0.01889	0.01969	0.01453

Table 4. Rotated Factor Pattern

Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
All Pickups, Mid & Large SUV, Vans	All Wagons	Mid & Small SUV	Comp & Mid Sedan	Ext Large & Large SUV	Mid Coupe & Sport	Small & Mid SUV	Mid & Large Sedan

Table 5. Important Loadings in Each Factor

A glance at the WTP's for the three BTS levels of the sedan segment reveals not much difference. To perform the CDA, instead of using all 144 WTP's for attribute dimensions simultaneously to estimate the coefficients, the stepwise selection method is used to select WTP dimensions which are statistically different across three groups of sedan type. The significant levels for entry and for stay are 0.01. The result of stepwise selection shows that only 60 out of 144 WTP dimensions are significant. Here, we had used a quadratic (as opposed to a linear) discriminant function because Box's M test, used to test the equality of variance-covariance matrices across all groups, indicated that the variance-covariance matrices are likely not equal (P-value < 0.001).

From the result of discriminant analysis, the first and second canonical correlations were found to be statistically significant as shown in Table 6. The first canonical variate separates compact sedan from mid-size and large sedans. The second canonical variate separates large sedan from compact and mid-size sedan, as shown in Figure 4.

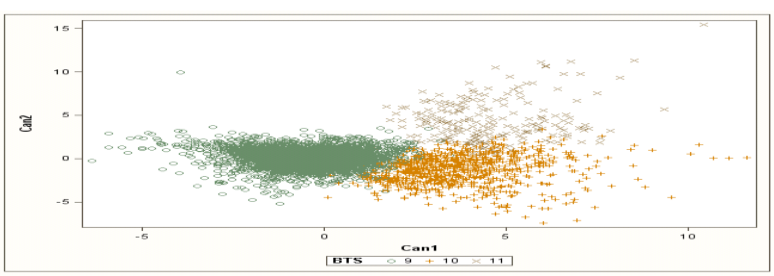


Figure 4. Scatter Plot of Canonical Variates 1 and 2

Multivariate Statistics and F Approximations					
S=2 M=28.5 N=4230.5					
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.19015252	182.41	120	16926	<.0001
Pillai's Trace	1.08437594	167.07	120	16928	<.0001
Hotelling-Lawley Trace	2.81520855	198.52	120	16383	<.0001
Roy's Greatest Root	2.14083144	302.00	60	8464	<.0001
NOTE: F Statistic for Roy's Greatest Root is an upper bound.					
NOTE: F Statistic for Wilks' Lambda is exact.					

	Canonical Correlation	Adjusted Canonical Correlation	Approximate Standard Error	Squared Canonical Correlation
1	0.825599	0.824236	0.003449	0.681613
2	0.634636	0.631507	0.006469	0.402763

Test of H0: The canonical correlations in the current row and all that follow are zero				
Likelihood Ratio	Approximate F Value	Num DF	Den DF	Pr > F
0.19015252	182.41	120	16926	<.0001
0.59723702	96.74	59	8464	<.0001

Table 6. Canonical Correlation and Test of Significance

To interpret the WTP dimensions that are most characteristic of group differences, the absolute of standardized canonical coefficients one and two are estimated. These coefficients capture the Eigenvalues of the distribution of the deviations from the nearest foreign cluster center. Then the coefficient values are ranked from the largest absolute values of the coefficients to the smallest absolute values. Table 7 and 8 present the ranks of top 10 dimensions of both canonical coefficients. The application of canonical discriminant analysis further validate our stand on using BT a primary classifier to differentiate amongst respondents.

Rank	WTP Dimensions (Standardized canonical coefficient 1)	Description
1	WTP_49	BTS: Compact Sedan
2	WTP_50	BTS: Mid-size Sedan
3	WTP_51	BTS: Large Sedan
4	WTP_133	Towing: 5000 lbs
5	WTP_127	Trunk Capacity: 21 cubic feet
6	WTP_63	BTS: Small 5-door
7	WTP_60	BTS: Full-size Van
8	WTP_134	Towing: 10,000 lbs
9	WTP_115	Safety Features: Antilock Brakes
10	WTP_64	BTS: Mid/Large 5-door

Table 7. Top 10 Dimensions of Absolute of Standardized Canonical Coefficient (1)

Rank	WTP Dimensions (Standardized canonical coefficient 2)	Description
1	WTP_50	BTS: Mid-size Sedan
2	WTP_51	BTS: Large Sedan
3	WTP_108	Number of Cylinders: 8 cylinders
4	WTP_133	Towing: 5000 lbs
5	WTP_127	Trunk Capacity: 21 cubic feet
6	WTP_105	Maximum Occupants: 8 occupants
7	WTP_106	Maximum Occupants: 12+ occupants
8	WTP_49	BTS: Compact Sedan
9	WTP_113	Roominess: Very spacious room
10	WTP_134	Towing: 10,000lbs

Table 8. Top 10 Dimensions of Absolute of Standardized Canonical Coefficient (2)

In the next subsection we look into techniques to reduce the length of the WTP vector to be used for developing time series models of customer preferences. Currently there are 144 elements in the WTP vector corresponding to various attribute levels as we explained in the introduction section. Looking into all 144 elements simultaneously would be rigorous and inefficient too as the number of parameters required in the models also increase correspondingly, hence we specifically look at elements within the WTP vector which are significant for utility computation and also show high levels of variation over time. Remaining attributes left out from the above analysis can be imputed using forecasts of the important attributes.

Dimensionality Reduction & WTP Vector Ordering

For reducing the overall length of the WTP vector to be used for developing time series models for each BT, we first need to rank attributes on two scales namely: 1. Significance of this attribute towards the overall utility of respondents belonging to the BT segment and 2. Variation observed in WTP for the attribute levels.

A possible approach to arrive at the importance of an attribute could be to use the slider data directly wherein respondents express their importance for various attributes. However as the exact nature of the transformation from the slider to WTP data is unknown and also since the slider data is heavily loaded with missing observations, we look directly into the WTP data for respondents for various attribute levels. It may be noted that where the importance in slider data is available for a respondent for an attribute, using the WTP data we would be obtaining rankings or the importance directly at the attribute level.

For quantifying the variance measure we use the SNR statistics. A big number for the importance and a large SNR value for an attribute level would indicate high predictability for the attribute level due to low variation. To obtain importance at attribute level we observe that greater the WTP for an attribute level for a respondent, greater would be its contribution towards the respondent's utility computation for a vehicle, i.e. to say a vehicle offering with greater values of the attribute level with respect to the market would correspond to higher utility level for the respective customer. Thus for each respondent we obtain the importance of the attribute level using the percentage or portion of each preferred WTP attribute level (maximum WTP level within each attribute) from the total WTP (sum across the maximum WTP levels within each attribute). Figure 5 shows the average percentage of each preferred WTP in Sedan segment.

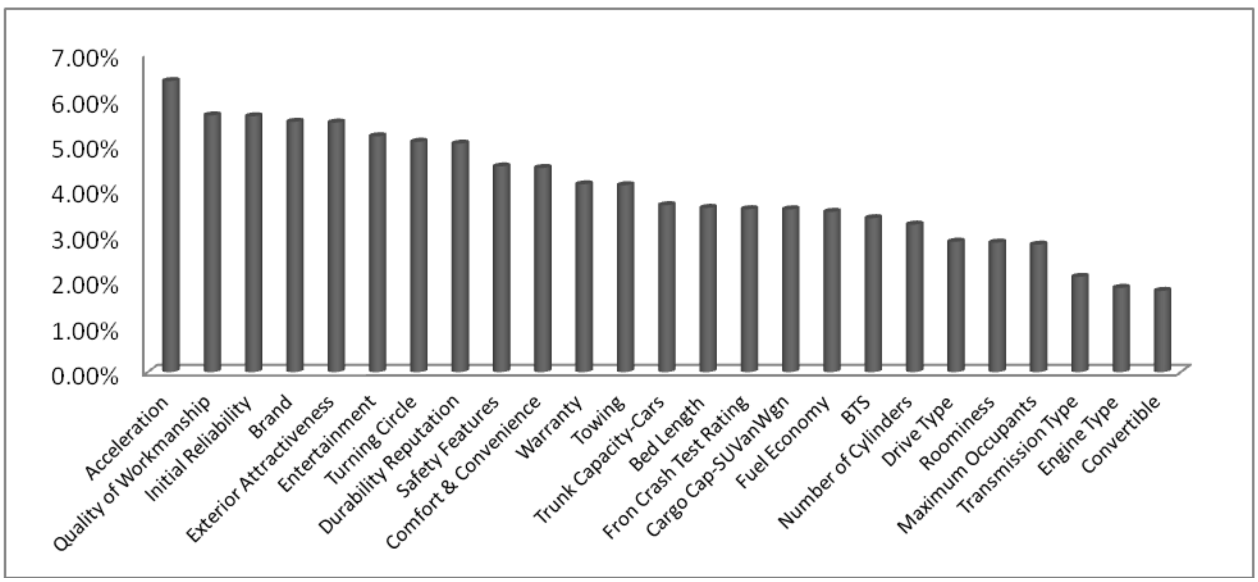
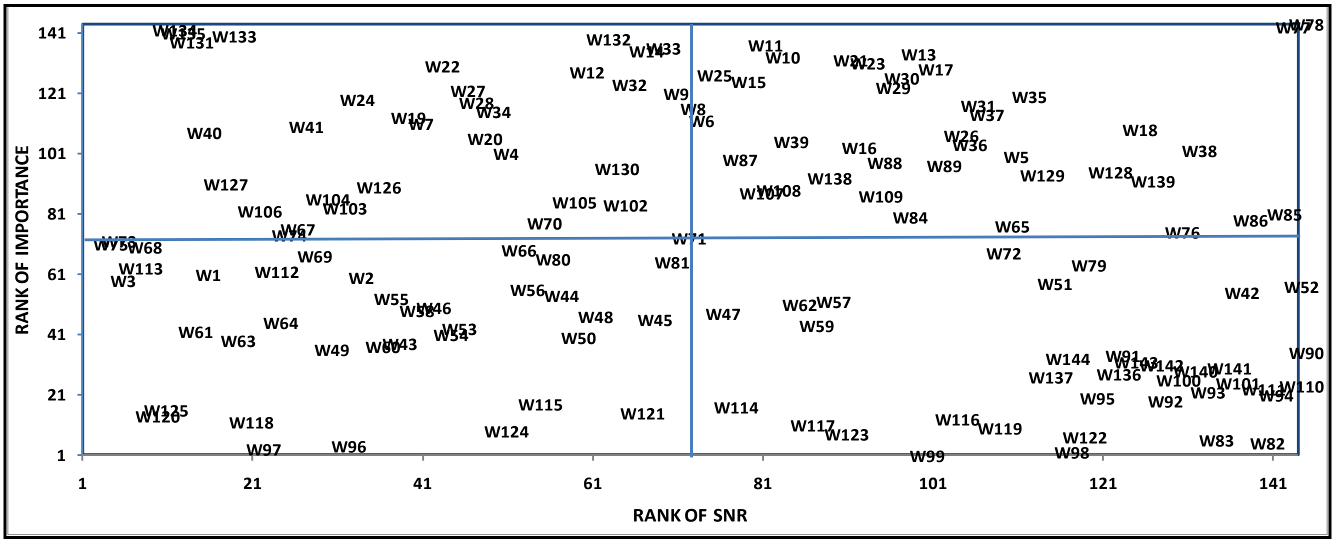


Figure 5. Percentage of Preferred WTP Levels

The SNR quantifier measures the ratio of energies of the signal and the noise components in every WTP level time-series data. Since the data tends to exhibit non-stationarity, a second order ARIMA model (the number of non-seasonal differences was 2) ARMA(p,2,q) was used to estimate energy components of the signal and noise. The values of p and q turned out to be 3 and 2 for most cases. From the results of SNR, the low-ranked 5 SNR (lowest values) are Exterior Attractiveness (better than most), Convertible (hard convertible top), Quality of Workmanship (better than most), Sport car and Convertible (soft convertible top). To find the important WTPs with low SNR values, quadrant analysis chart is shown in Figure 6. The vertical axis of this chart is the rank of importance of attributes using the average importance rating by respondents. Since the average importance ratings are for attributes only (not attribute levels), the average of WTPs within an attribute are used to rank WTP levels within an attribute. The horizontal axis of the chart is the rank of SNR values by 144 WTP attribute levels. The most important WTPs with low SNR values are in the bottom right of the quadrant analysis chart. Some of the WTPs in the bottom right of the quadrant analysis chart include Durability Reputation (better than most), Durability Reputation (among the best), Front Crash Test Rating (rated 4 stars), Quality of Workmanship (better than most) and Quality of Workmanship (among the best). These WTP attributes are significantly important and yet difficult to predict. These factors will be considered for the application of the unconventional approaches to be taken to develop the prediction methodology of customer preferences. Thirty-two out of 144 WTPs are in this area of the quadrant chart.



SUMMARY

This paper presents an analysis of customer preference of automobile products in terms of Willingness-To-Pay (WTP) of vehicle attributes using SAS®, including customer segmentation, and dimensionality reduction of WTP vector. This analysis addresses issues related to the preparation of customer inputs for effective prediction of customer preferences. A gist of our findings is presented below.

Multicolinearity Reduction**:** Factor analysis of the WTP data is performed to identify the salient attribute space with which to segment the respondent data. In specific, PCA (Principal Component Analysis) was used to identify attributes exhibiting maximum variability amongst customers and also eliminate some redundant attributes for subsequent analysis. The effect of price is not considered in our analysis, i.e. as WTP for attributes specifies a dollar amount associated with per unit of the attribute one would intuitively expect attributes like e.g. comfort & convenience features to show max variability in WTP across customers. As this can happen, attributes which ride heavily on price post facto are eliminated. PCA analysis thus reveals BTS to be the most variable attribute dimension across customers.

Customer Segment Classification: Segmentation of customers was performed on the reduced attribute space (devoid of high levels of multicollinearity) that has resulted from the application of PCA. Potential customers are classified by their Body Type preferences. EFA (Exploratory Factor Analysis) using the BTS attributes further reveals that the canonical dimensions correspond to BT segmentation. Current focus is on Sedan and SUV segments.

Dimensionality Reduction & WTP Vector Ordering**:** The customer segmentation step helps us in looking at aggregate preferences of a particular customer segment. Given the nature of the WTP data as described earlier, even within these segments there are 144 attribute levels to be modeled for. This step aims to rank attributes levels according to their importance and predictability. Importance ranking is derived based on how much the customers consider the specified attribute to make their product decisions. A signal-to-noise ratio (SNR) measure is derived from the customer data to rank attribute levels in terms of their predictability, as higher the SNR larger will be the signal than noise indicating higher predictability and vice versa. The analysis will allow a significant reduction in the dimension of the state vectors to be considered for forecasting customer preference shares. Further attributes based on values of these two metrics would be modeled for separately. These attributes are identified for the sedan and SUV segments.

In conclusion, our analysis shows that respondents show maximum variability in specifying their WTP for Body Type (BT) attribute and hence classification based on BT associations is the most suitable way to differentiate between respondents from available data. In case of dimensionality reduction of WTP vector, we propose a twofold criteria for WTP rankings. The first addresses the importance of WTP for an attribute level from a respondent's utility maximization standpoint. The second criteria addresses the predictability of WTP from the available data using

Signal-to-Noise Ratio (SNR). The combination of both criteria provides excellent insights to the analyst regarding importance and predictability of WTP data, and also provides a tractable way to the dynamic modeling of customer preferences.

REFERENCES

1. Krishnan, V. and K.T. Ulrich, *Product Development Decisions: A Review of the Literature.* Management Science, 2001. **47**(1): p. 1-21.
2. Louviere, J.J. and T. Islam, *A comparison of importance weights and willingness-to-pay measures derived from choice-based conjoint, constant sum scales and best–worst scaling.* Journal of Business Research, 2008. **61**(9): p. 903-911.
3. Garcia, P.M., *Measuring Willingness-To-Pay in Discrete Choice Models with Semi-Parametric Techniques.* Economics and Econometrics Research Institute (EERI), 2005.
4. Breidert, C., *Estimation of willingness-to-pay. Theory, measurement, and application.* Doctoral thesis, WU Vienna University of Economics and Business, 2005.

ACKNOWLEDGMENTS

This research is supported by General Motors-India Science Lab (GM-ISL).

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Name: Akkarapol Sa-ngasoongsong
Enterprise: Oklahoma State University
Address: 322 Engineering North
City, State ZIP: Stillwater, Oklahoma, 74075, USA
E-mail: akkarap@okstate.edu

Name: Satish T.S. Bukkapatnam
Enterprise: Oklahoma State University
Address: 322 Engineering North
City, State ZIP: Stillwater, Oklahoma, 74075, USA
E-mail: akkarap@okstate.edu

APPENDIX A: WTP ATTRIBUTES AND LEVELS

1. Acceleration*
   - **1. Somewhat slow accel** (1)
   - 2. Somewhat fast & smooth accel (1-2)
   - 3. Fast & smooth accel (2-3)
   - 4. Very fast & smooth accel (3-4)

2. Bed Length
   - **1. 5 feet** (5)
   - 2. 6.5 feet (4-6)
   - 3. 8 feet (5-7)

3. Brand
   - 1. Acura (6 -8)
   - 2. Audi (7 -9)
   - 3. BMW (8 -10)
   - 4. Buick (9 -11)
   - 5. Cadillac (10-12)
   - 6. Chevrolet (11-13)
   - 7. Chrysler (12-14)
   - 8. Dodge (13-15)
   - 9. Ford (14-16)
   - 10. GMC (15-17)
   - 11. Honda (16-18)
   - 12. HUMMER (17-19)
   - 13. Hyundai (18-20)
   - 14. Infiniti (19-21)
   - 15. Isuzu (20-22)
   - 16. Jaguar (21-23)
   - 17. Jeep (22-24)
   - **18. Kia** (25)
   - 19. Land Rover (23-26)
   - 20. Lexus (24-27)
   - 21. Lincoln (25-28)
   - 22. Mazda (26-29)
   - 23. Mercedes (27-30)
   - 24. Mercury (28-31)
   - 25. MINI (29-32)
   - 26. Mitsubishi (30-33)
   - 27. Nissan (31-34)
   - 28. Pontiac (32-35)
   - 29. Porsche (33-36)
   - 30. SAAB (34-37)
   - 31. Saturn (35-38)
   - 32. Scion (36-39)
   - 33. Subaru (37-40)
   - 34. Suzuki (38-41)
   - 35. Toyota (39-42)
   - 36. Volkswagen (40-43)
   - 37. Volvo (41-44)

4. Body Type & Size
   - **1. Compact Coupe** (45)
   - 2. Mid-size Coupe (42-46)
   - 3. Mid-size Ext Pickup (43-47)
   - 4. Large Ext Pickup (44-48)
   - 5. Mid-size Crew Pickup(45-49)
   - 6. Large Crew Pickup (46-50)
   - 7. Mid-size Std Pickup (47-51)
   - 8. Large Std Pickup (48-52)
   - 9. Compact Sedan (49-53)
   - 10. Mid-size Sedan (50-54)
   - 11. Large Sedan (51-55)
   - 12. Sport Car (52-56)
   - 13. Small SUV (2dr) (53-57)
   - 14. Mid-size SUV (2dr)(54-58)
   - 15. Small SUV (4dr) (55-59)
   - 16. Mid-size SUV (4dr)(56-60)
   - 17. Large SUV (4dr) (57-61)
   - 18. Extra Large SUV (4dr)(58-62)
   - 19. Mini-van (59-63)
   - 20. Full-size Van (60-64)
   - 21. Small Wagon (61-65)
   - 22. Mid-size Wagon (62-66)
   - 23. Small 5-Door (63-67)
   - 24. Mid/Large 5-Door (64-68)

5. Comfort & Convenience
   - 1. Adjustable pedals (65-69)
   - 2. Front heated seats (66-70)
   - 3. Keyless entry (67-71)
   - 4. Leather seats (68-72)
   - 5. Maint interval indicator (69-73)
   - 6. Navigation aid (70-74)
   - 7. Remote starter (71-75)
   - 8. Separate climate cntrls (72-76)
   - 9. Steering wheel cntrls(73-77)
   - 10. Sunroof Moonroof (74-78)
   - 11. Telescopic steering (75-79)
   - 12. Wireless connectivity (76-80)

6-17. Convertible
   - **1. No convertible top** (81)
   - 2. Soft convertible top (77-82)
   - 3. Hard convertible top(78-83)

7-18. Drive Type
   - **1. FWD** (84)
   - 2. RWD (79-85)
   - 3. 4WD (80-86)
   - 4. AWD (81-87)

8-19. Durability Reputation*
   - **1. Does not stand out** (88)
   - 2. Better than most (82-89)
   - 3. Among the best (83-90)

9-20. Engine Type
   - 1. Gasoline (84-91)
   - **2. Diesel** (92)
   - 3. Flexible fuel (E85) (85-93)
   - 4. Hybrid (86-94)

10-21. Entertainment
   - 1. CD changer (87-95)
   - 2. DVD player (88-96)
   - 3. Satellite radio (89-97)

11-24. Exterior Attractiveness*
   - **1. Does not stand out** (98)
   - 2. Better than most (90-99)
   - 3. Among the best (91-100)

12-25. Front Crash Test Rating
   - 1. Not rated (92-101)
   - **2. Rated 1 or 2 stars** (102)
   - 3. Rated 3 stars (93-103)
   - 4. Rated 4 stars (94-104)
   - 5. Rated 5 stars (95-105)

13-26. Fuel Economy*
   - **1. 15 miles per gallon** (106)
   - 2. 25 miles per gallon (96-107)
   - 3. 35 miles per gallon (97-108)
   - 4. 45 miles per gallon (98-109)
   - 5. 55 miles per gallon (99-110)

14-27. Initial Reliability*
   - **1. Does not stand out** (111)
   - 2. Better than most (100-112)
   - 3. Among the best (101-113)

15-28. Maximum Occupants
   - **1. 2 occupants** (114)
   - 2. 4 occupants (102-115)
   - 3. 5 occupants (103-116)
   - 4. 7 occupants (104-117)
   - 5. 8 occupants (105-118)
   - 6. 12+ occupants (106-119)

16-29. Number of Cylinders
   - **1. 4 cylinders** (120)
   - 2. 6 cylinders (107-121)
   - 3. 8 cylinders (108-122)
   - 4. 10 cylinders (109-123)

17-30. Quality of Workmanship*
   - **1. Does not stand out** (124)
   - 2. Better than most (110-125)
   - 3. Among the best (111-126)

18-31. Roominess
   - **1. Compact roominess**(127)
   - 2. Spacious roominess (112-128)
   - 3. Very spacious room (113-129)

19-32. Safety Features
   - 1. Accident alert system (114-130)
   - 2. Antilock brakes (115-131)
   - 3. Daytime running lights (116-132)
   - 4. Electronic brake assist (117-133)
   - 5. Front side airbags (118-134)
   - 6. Hands-free phone (119-135)
   - 7. Rear object detection (120-136)
   - 8. Run-flat tires (121-137)
   - 9. 2nd row side airbags (122-138)
   - 10. Side curtain airbags (123-139)
   - 11. Stability control (124-140)
   - 12. Theft tracking (125-141)

20-44. Trunk Capacity-Cars*
   - **1. 5 cubic feet** (142)
   - 2. 13 cubic feet (126-143)
   - 3. 21 cubic feet (127-144)

21-45. Cargo Cap.-SUVanWgn*
   - **1. 12 cubic feet** (145)
   - 2. 30 cubic feet (128-146)
   - 3. 48 cubic feet (129-147)
   - 4. 66 cubic feet (130-148)

22-46. Towing*
   - **1. 0 pounds** (149)
   - 2. 1,000 pounds (131-150)
   - 3. 2,500 pounds (132-151)
   - 4. 5,000 pounds (133-152)
   - 5. 10,000 pounds (134-153)
   - 6. 15,000 pounds (135-154)

23-47. Transmission Type
   - **1. Manual transmission** (155)
   - 2. Automatic trans (136-156)
   - 3. Auto + manual override (137-157)

24-48. Turning Circle*
   - **1. Large** (158)
   - 2. Intermediate (138-159)
   - 3. Small (139-160)

25-49. Warranty (basic/pwrtrn)
   - **1. 3-year /3-year** (161)
   - 2. 3-year/5-year (140-162)
   - 3. 3-year/7-year (141-163)
   - 4. 4-year/4-year (142-164)
   - 5. 5-year/10year (143-165)
   - 6. 10year/10year (144-166)