Agent Based Model and Statistical Engineering of Automobile Purchase Behavior and Retail Actions

James G. Wendelberger
Urban Science Applications, Inc., 620 Los Pueblos, Los Alamos, NM 87544

Abstract
For several decades the automobile purchase behavior of new cars has been modeled and utilized by automobile manufactures for business decisions ranging from retail location, retail operations, consumer behavior, and competitive actions and reactions. Up until now these models have been optimized on an individual or ad hoc basis. It is proposed that these models are actually interrelated and should be developed using a statistical engineering process. This process is demonstrated with an agent based proto type. The prototype displays unit sales and outlet profitability based upon outlet attractiveness and competitive actions and reactions on a daily basis. This prototype will serve as the base from which an all-encompassing state-of-the-art automotive market model is planned to be developed through a statistical engineering process. The entire model and real world data will adhere to the principals of the scientific method.

Key Words: Agent based model, statistical engineering, automobile, purchase behavior, retail actions, scientific method

1. Mathematical and Statistical Models in the Automobile Industry

The main goal of this undertaking is to infuse science into the automobile service marketplace. Ways to accomplish this, using the scientific method, SM-14, Edmund (2012), are provided in Lampathaki et al (2012). The idea of using the consumer or the new vehicle as the "atom" for the modeling analysis follows the physical process described in Stringer (2008). There are many mathematical and statistical models utilized in the automobile industry. Examples of a very small subset of these models are exhibited by the report titles: Lower variance by minute modeling, Distribution to lower geographic level, Extraction of fleet and retail, US Markets and health during the oil crisis, Disaster Forecasting: New Orleans and The Netherlands – Expert Opinion, Objective versus subjective, Behavior versus perception, Forecasting automobile sales – is there other data that can help? Or, is the history the best?, The role of "novelty detection" here., Classification of data types and statistical models for solutions to business issues., Internal and External Model Validity, Dedistortion, Distribution, Examples – all enterprise Analyses to date, Include National Analysis of Markets as Time Series, Sample Size Program Help File sample size program, Statistical Memorandum - Manufacturer Op per Dealer Outline, The Use of Location in Telemarketing, Campaigns to Measure and Optimize Response Probability, A Trend-Cycle Forecasting Model, Point Prediction for a Logit Regression Model, Segmentation Issues, A Non-Linear, Probability Based Approach to Prospect List Selection, Outlet Characteristic Analysis - Micro-Location Parameter Estimation and Significance, Profitability Through Lifetime Value Modeling and Analysis - The Insurance Company, Lifetime Value Modeling and Analysis - The Bank, Location and Geospatial Data in Statistical Models for Better Business Decisions, Integration of Models for Predictive Modeling, A Worldwide Automotive Registration Database with Analysis on Demand, National Center for Database Marketing - Chicago 1995 Conference Report, Interpretation of Regression Estimates, Program Plus Sales - Statistical Significance and Estimation, The Statistical Modeling Process, Market Ranking From Demographic Factors, Examples of Linear Regression Statistical Significance, Examples of Linear Regression Outliers, Examples of Linear Regression R-Squared, Customer Satisfaction Index Sign Test Example, Finding Control Markets, Practical Experimental Design under Multiple Prior Information Sources, Regression on Sampled Signals and Curves - A P-Spline Approach, Percentiles - An Example of Size 330, Future Network Planning, Non-stationary spatial modeling using a process convolution approach, Robust Modeling, A Partial Order Approach to Record linkage, Aberrant Automotive Retail Registration Data Detection, Different Means, Data Checks for Two Year Automotive Registration Data By Geographic Unit, Statistical Terms Matrix, Statistical Analysis Skill Definition, Saab Fleet Redistribution, 2004 - 2005 Worldwide Analysis Summary, Computer Component Configuration Price Estimation, Statistical Matrix Definitions, Mixtures at the Interface, Trend-Cycle Forecasting of Sales Expected , Registrations and Industry Registrations, Possible Micro-Location Factors, Excel Macro for Trend Cycle Model – NLIN, Dealer Key Performance Internet Pilot , Research Effectiveness vs Standard, A Trend Cycle Forecast Model of Automotive Sales, Lean Six Sigma at A Company, Valid Transformations for Regression Variables and Sample Size Needed for Assessing Compliance to a Standard. It is not expected that the titles sufficiently describe the various models, however, the titles should provide an indication of the model variety. In what follows all of these models and many others are included in the “overall model”.

1.1 Overall Model Technical Description

The capabilities and requirements of the overall model are defined based upon the following top level entities: manufacturers, new vehicles, sellers and buyers. The manufacturers create the new vehicles which are sold by the sellers to the buyers. This process is called the new vehicles sales flow process and is illustrated in Figure 1. The next level of the model further refines the process definition. The manufacturers are often divided into a “client” manufacturer (or manufacturers) and the competitive manufacturers. The new vehicles may be divided into those that compete with one another versus those that do not compete with one another. The competing vehicles comprise “segments”. Segments are used by all of the manufacturers for various purposes including: market share, product planning, advertising, sales and marketing. The sellers include the brick and mortar stores which are also called automotive dealerships. The dealerships may be further subdivided into urban and rural or in an auto row versus isolated, etc.

Manufacturers with internet presence or having access to internet visit and search data often utilize a set of “leads” which may be used by the marketing groups to define motivated or interested prospects. Another source of buyers includes manufacturer employees or their suppliers who may buy through manufacturer incentive/loyalty employee plans. The new vehicles are sold to the buyers. Buyers may be the standard retail buyer, an internet buyer or a corporate entity buying for business purposes. Among many, examples of corporate buyers include: taxi companies, rental cars companies, police departments and school systems. The modeling process utilizes the scientific method to determine the theoretical structure of this process. This theoretical structure may then be utilized for various purposes including: forecasting, network change, incentive planning, product planning, manufacturing optimization, marketing, advertising, etc.

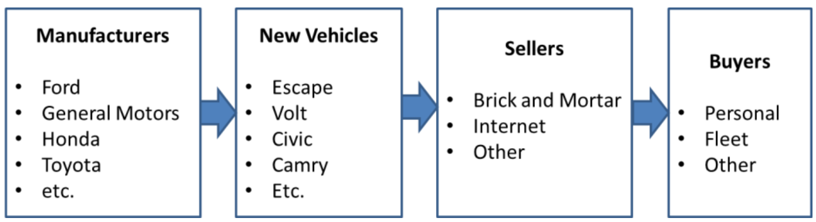

Figure 1: Overview of the Overall Model of the Real World Flow Process

Included in Figure 1 are the rightward pointing arrows which represent new vehicle design and assembly of vehicles from the raw materials, completed product shipped to the sellers, and the new vehicles sold from the sellers and delivered to the buyers. The most common path represents the retail buyer process. There are also implicit processes required to accomplish the overall model flow process. These may include the buyer visiting a seller location, a buyer searching online for a vehicle, a manufacturer creating a “green” product, a seller requesting specific new vehicle types for inventory, etc. In effect these implicit processes show that there is in fact information flow in both directions. These considerations result in the twelve process flow arrows in the overview of the overall model of the real world automotive process in Figure 2.

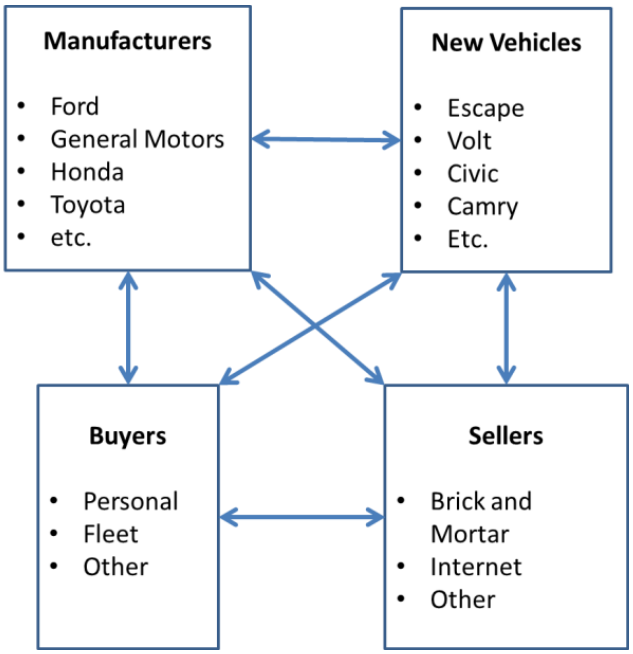

Figure 2: Overview of the Overall Model of the Real World Automotive Process

The processes represented in Figure 2 are used to illustrate the complexities which may be accounted for in a theoretical model. Further complexities which are implicit in this representation include: product shortages or recalls, manufacturer labor issues, manufacturer incentives, product quality issues, seller loyalty and integrity issues, seller locations, seller inventory and marketing, buyer locations, and underlying data relating to all of these issues. All of these issues may contribute to the theoretical model of the process depicted in Figures 1 and 2. Some of these complexities result in feedback shown by the reverse pointing and additional arrows in the Figure 2. There are four boxes in the figure which represent organizational or group behavior. These along with the twelve arrowheads in Figure 2 represent the following groups (1-4) or flows (5-16):

1. Manufacturers
2. New Vehicles
3. Sellers
4. Buyers
5. Manufacturer to New Vehicles
6. Manufacturer to Sellers
7. Manufacturers to Buyers
8. New Vehicles to Manufacturers
9. New Vehicles to Sellers
10. New vehicles to Buyers
11. Sellers to Manufacturers
12. Sellers to New Vehicles
13. Sellers to Buyers

14. Buyers to Manufacturers
15. Buyers to New Vehicles
16. Buyers to Sellers

One may consider outside factor influences on the 16 categories as included in each individual model.

1.2 Fundamental Processes

These 16 categories represent fundamental processes within the automotive industry. It is these fundamental processes which are modeled in order to accomplish the goals within the automotive business community. Below are examples from each category:

1. The manufacturer decides what to build, how to incentivize, and how to warranty its brand(s).
2. The new vehicles are built (or await custom ordering) and form a competitive group of transportation options.
3. The sellers form a retail network for each brand across the country.
4. The buyers consider a new vehicle for purchase.
5. The manufacturer may increase or decrease production, decide where to build the vehicles and transport the raw materials to the plant(s).
6. The manufacturers decide who will sell their products and from where they will sell them. This is dealer network planning at the country level.
7. The manufacturers may advertise and provide incentives to buyers.
8. New vehicles may have labor issues at a plant or suffer from part shortages or recalls or quality issues (both good and bad) which influence actions required of the manufacturer.
9. New vehicles may provide sellers with warranty work and regular service and maintenance.
10. New vehicles may be desired by a buyer irrespective of the seller/delivery options.
11. A potential seller may request a new dealer outlet location from a manufacturer. This is also governed by franchise laws.
12. Sellers may request inventory or sold vehicles from the new vehicles.
13. Sellers sell cars to buyers and may attract buyers via test drives, price or availability.
14. A buyer may provide direct quality and experience feedback to the manufacturer.
15. Buyers may as a group demand more “green” or sport utility vehicles be made.
16. Buyers may show up at a particular seller location.

We see that the automotive market is vast with many opportunities for optimization and revenue.

2. Examples

2.1 Organizational or Group Behavior

It is the vast nature of the business that has hampered efforts to document and categorize the models utilized in the industry. That is reflected by the multitude of examples which may be provided for these 16 fundamental categories.

2.1.1 Manufacturers

Consider the "Manufacturers" box from Figure 2. This box represents all of the actions that may be taken by a manufacturer and the competitive manufacturers. These include:

1. Dealer network planning
2. Develop new vehicle models
3. Create dealer agreements
4. Advertise and Advertising strategies
5. Marketing - Incentives
6. Profitability
7. Automotive Research

Manufacturers make many decisions related to the automobile business. These may include: whether or not to do data analysis internally, negotiations with unions, determining various budgets, such as, advertising, marketing, network planning, manufacturing, creating dealer agreements, types of vehicles to manufacture, automobile research, conformance to government mandates, incentive, inventory, manufacturing quotas, etc. For the new vehicle process described here we will consider the issues related to dealer network planning. These may include: dealer location issues, market segmentation decisions and incentive planning.

2.1.2 New Vehicles

New vehicles box in Figure 2 represent a collection of new vehicles that currently exist in the marketplace. New vehicle issues include whether a manufacturer properly represented in all vehicle segments, which vehicles are entering the marketplace, which vehicles are departing the marketplace and which vehicles are staying in the marketplace?

2.1.3 The Sellers

Sellers are represented in the Sellers box of Figure 2. Sellers may include the dealer networks and in some countries direct purchase from manufacturers, often via the internet. In addition, it is possible to sell a vehicle across country borders.

2.1.4 The Buyers

Buyers are represented in the Buyers box of Figure 2. There are two main categories of buyers: retail and fleet. Retail buyers are usually individuals or household members. Fleet buyers include businesses such as taxi companies, delivery services and corporate vehicles. The exact definition of these categories varies by country. For example, in some parts of Europe a corporate vehicle may be purchased by the employing company and yet considered as a retail purchase would be used in the United States. There are also other vehicles, such as, those coming into a country through its borders and those from manufacturer incentive programs for its employees.

2.2 Flows between Organizational or Group Behaviors

Twelve of the fundamental categories are flows between the four different fundamental categories.

2.2.1 The Manufacturers Flow to New Vehicles

The normal flow is to direct and pay for the manufacture of new vehicles which are then provided to the new vehicle marketplace. This will include getting materials and parts to the manufacturing facilities, assembly and completed vehicles.

2.2.2 The Manufacturers Flow to Sellers

The Manufacturers Flow to Sellers will include facility requirements, dealer agreements, inventory directions and others.

2.2.3 The Manufacturers Flow to Buyers

The buyers tell us about themselves through various mechanisms. As with most things there are many ways they do this. The most common way is by their name. Another one of the most utilized ways is the address to which they register their new automobile. Using these two pieces of information along with when a vehicle was registered may be used to determine the likelihood of these purchases being personal versus a fleet customer.

The address may be used to determine an exact location of a customer. The address may represent a person's home or place of business or some other address. In the business we call this the location of the customer and rarely determine whether or not the address is actually a home, business or other place. For example, the address is sometimes a harbor where a boat is registered. Presumably, a customer with a boat registered in a harbor could in fact live almost anywhere. One of the products produced by the addresses is an exact dot map. In this map the exact locations of geocoded addresses are placed on a map of some geographical area. The density of the dots on the map represents the preponderance of the buyers' locations.

Even in this very simple example of the use of the registration data one uncovers many issues that must be considered. These include:

1. What name is provided?
2. What address is provided?
3. Is the name the same, written differently or misspelled when looking for multiple purchases to the same person or entity?
4. Is the address the same, written differently and/or misspelled when looking for multiple purchases to the same person or entity?
5. Were all addresses geocoded or are some non-decodable?

6. To what different levels may the addresses be geocoded? Some may be to zip others to street, etc.
7. What do we do with addresses that are not geocoded?
8. What rules do we use to determine personal versus fleet and other registrations?

Another product which may be formed from the locations is a random dot map. In this map the addresses are geocoded to some area of geography. One of the most common is zip code. Then the number of customer's within each zip code is counted. A random number of dots are placed in the zip code in proportion to this counted number.

This random dot map is similar to the exact dot map. It has advantages and disadvantages. The advantage is that the eye is more likely to properly perceive the dot density. The disadvantage is that the dots may be placed in, for example, uninhabitable places. The random dot map also has the advantage that through the proportionality constant or scaling it may be used to display maps that would otherwise just be a smear of ink due to too many dots on the map. One sure giveaway to whether you are looking at an exact dot map or a random dot map is the small scale order on the map. On the exact dot map one can perceive boundaries and curves which represent for example, parks and streets, whereas on the random dot map there is very little fine structure visible.

2.2.4 The Sellers Flow to Buyers.

The sellers provide the buyers with new vehicles. They may also supply them with financial assistance in order to complete the purchase.

2.2.5 The Buyers Flow to Sellers

The buyers provide money to the sellers. They may draw this money from personal monetary sources or personal financing. They may make only a partial payment (down payment) and then the vehicle may be used for collateral on the balance.

2.2.6 Other Flows

The New Vehicles Flow to Manufacturers, The New Vehicles Flow to Sellers, The New Vehicles Flow to Buyers, The Sellers Flow to Manufacturers, The Sellers Flow to New Vehicles, The Buyers Flow to Manufacturers and The Byers Flow to New Vehicles.

2.2.7 More Complex Models

Of course there are problems which require multiple categories and multiple flows. This complicates the categorization as there are 4 possibilities for 4 categories, 1 possibility for all for categories and then 6 possibilities for 2 flows, 4 possibilities for 3 flows and then multiple combinations of flows and categories. These multiple categories and flows will be treated as consisting of combinations of the basic categories and flows (called fundamental categories) outlined in Figure 2. In the business world these multiple flow situations are usually treated by working with the fundamental processes.

3. More Expansive Real World Transportation Related Processes

There are related processes which are broader than the processes portrayed in Figure 2. These include more expansive processes. This is portrayed in Figure 3. These contain, as a subset, the automobile processes listed above.

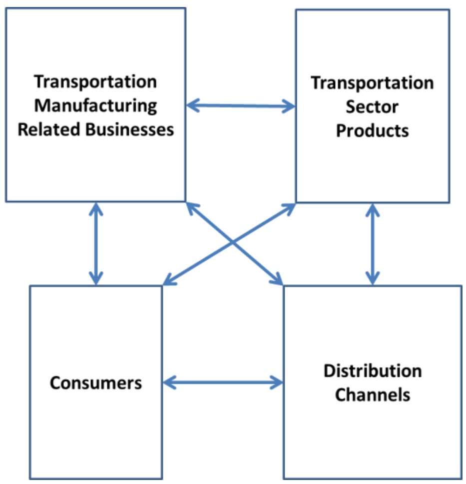

Figure3: More Expansive Real World Transportation Related Processes

The more expansive real world transportation related processes which can benefit from statistical modeling include:

1. Manufacturers of motorcycles, boats, motor homes, buses, trucks, airplanes, trains, etc.
2. Financial aspects of the automobile manufacturers including: incentives, vehicle financing, dealer incentive programs, leasing, etc.
3. Financial aspects of transportation.
4. Transportation products: new and used motorcycles, boats, motor homes, buses, trucks, airplanes, trains, etc. and used automobiles.
5. The distribution channels include: used car dealers, off lease new vehicle auctions, internet, brick and mortar, business to business, consumer to consumer, private sale, etc.
6. Consumers include anyone who uses the transportation system.
7. Interactions between the transportation related businesses, transportation sector products, distribution channels and consumers.

As one might imagine the possibilities for the transportation related process solutions are huge.

4. Technical Requirements of the Model

The technical requirements of the model are:

1. Faithfully Mimics Reality
2. Useful for the Identified Purpose

To insure that our overall model mimics reality we include expert knowledge of the entire process. This knowledge is then vetted via case studies and data in model checking techniques. To insure that the model is useful for the identified purpose the model is tested and utilized in environments as close as possible to the identified purpose. All of this is done under state-of-the-art scientific model building and testing methodologies.

5. Available Data:

One of the hardest aspects of this methodology is acquiring useful data for the identified process. The next hardest aspect is to incorporate the appropriate model with or without available data. Data available has changed over time from a few decades ago until today. It is also still only minimally available in some countries around the world.

Data that is available for the various categories in Figures 3-17 may be classified using the first letter of the individual categories and the two letters of the directed flows. This categorization will be shorthand and label the categories and flows as: M, N, S, B, MN, MS, MB, NM, NS, NB, BM, BN, BS, SM, SN, and SB. We can then extend this labeling to incorporate various automobile systems. We do this by adjoining the appropriate categories by a hyphen. For example, if we wanted to look at the number of sellers (dealers) and the number of buyers we would be considering a S-B problem. If instead we are concerned with how vehicles move from sellers to buyers we would be considering either an SB problem or an S-SB-B problem. Which one depends upon whether or not we are only concerned about flow or in addition something from categories S and B too. We next append the subscript d to indicate that we are talking about data for the specific problem. For example, S-SB-B_d means that we have data for the S-SB-B problem. Presumably, this would include data about S, B and the flow SB. The subscript m indicates that we are concerned about a model for certain data. For example, S-SB-B_m means that we have a model for the S-SB-B_d data and the S-SB-B problem. And finally, we may include a superscript T to indicate that we are considering the broader case of transportation related data or models, as for example, S-SB-B_m^T means that we have a model for the S-SB-B_d^T data and the S-SB-B^T problem.

We now have shorthand for our problem, data, and model solutions. For example, the automotive process depicted in Figure 1 is denoted: M-MN-N-NS-S-SB-B, its associated data is $\text{M-MN-N-NS-S-SB-B}_d$ and models describing this process may be written $\text{M-MN-N-NS-S-SB-B}_m$.

Using this notation it becomes quite apparent why keeping track of all the modeling is such a hard problem. And this notation does not even specify the specific problems, available data, statistical models and overall solution.

Let us next look at the dealer network planning problem. This is minimally a S-B problem. That is we ask where to place the sellers given the distribution of buyers.

To make this all work we are going to classify data and statistical models too. This will involve input and possible outputs of statistical models. Also included will be assumptions or constraints. For example, the number of sales is non-negative or the dealer location is not in a lake, etc.

What arguably is the most useful data in the new automobile industry, Figure 2, is the buyer data. This is where the money starts from and flows back up, sellers getting some share, new vehicles having capital for manufacturer and the manufacturers themselves getting the balance.

The best data available for our purposes is the location of actual buyers in a market. This represents buyer behavior rather than buyer perceptions. For this reason it is imperative that we utilize buyer sales or registration records as a primary source of data. In some markets this data does not exist. In this case and only in this case, we then rely on surrogate data. The surrogate data is used to estimate the registration or sales.

The data on new vehicles purchased is divided into two main categories: registrations and sales. Registration data is that data which is required by some governmental agency. This data is sent to the governing body by the seller. The data, in electronic form, is then sold or provided by the governmental agency. The sales data is data collected by the seller or dealership. The sales data may or may not be provided to the manufacturer. The sales data is may be aggregated and provided for analytical purposes.

6. Information

Aside from data we sometimes utilize: expert knowledge, expert opinion, constraint enforcement, client restrictions, approximations, etc. in our models.

7. Available Models/Knowledge

Many types of data described above have been used in the automotive marketplace modeling processes. The models and knowledge that result from this experience are incorporated into the overall or theoretical comprehensive model. The most powerful relationship detected between buyer and seller is location, as supported by the phrase "location, location, location".

8. Model Philosophy

The main philosophy is to document the models and data used for specific purposes. This will include a priority setting of data and models as higher quality data and models become more and more available. The model is structured to use minimal to exhaustive data and everything in between.

Models correspond in Figure 1 to the boxes labeled manufacturer, product, seller and buyer and to one each to each direction of the three arrowheads. The arrows correspond to manufacturer influencing product, product influencing manufacturer, product influencing seller, seller influencing product, seller influencing buyer and buyer influencing seller. This representation is not meant to imply there is no product to buyer influence but rather to represent that influence by the product influencing seller and then the seller influencing the buyer. In this way we hope to simplify at least the Figure 1 representation of the real world process.

For example, a manufacturer may advertise a product on a national level thus reaching the buyer directly. This is represented in the figure by the three rightward pointing arrowheads. In a similar way the buyer/prospect may respond directly to the manufacturer through a webpage or survey. This contact would be represented by the three leftward pointing arrowheads.

The model is listed in function form to reveal the likelihood of purchase of a specific product from a specific seller to a specific buyer. This likelihood is a function of seller attractiveness, product desirability and buyer desire. The main model plays off the seller attractiveness product desirability and buyer desires to determine this probability. This model may be written in mathematical notation as:

Probability(i,j,k) = attractiveness (i,j,k)/ sum attractiveness (I,J,K)

where i refers to the seller, j to the vehicle category and k to the geographic location of the buyer. Capitalized letter refer to the total of each category.

Many of the models created here are based upon geographic areas, time period, automobile models, and selling dealer. These categories should be subdivided to a fine level of detail. For example, census block groups, daily or hourly sales, automobile make/model with specific options and selling dealer. The model is then formulated at this fine level and then aggregated to supply the model deliverables. These deliverables might include: market area such as state or country, monthly or yearly, automobile segment, such as luxury car, and brand of selling dealer. The unbiased estimation at the fine level of detail and then aggregation to a higher level will reduce the internal model variability or variance of the estimates.

It is tempting to believe that the demographics or economic factors are a better predictor of automotive sales or registrations than the sales or registrations themselves. This is a fallacy and should be guarded against at all times. The registrations and sale implicitly incorporate not only the demographic and economic factors available for analysis but also all other demographic and economic data, otherwise, unavailable for analysis. This has been one of the success stories of Urban Science. Utilize the sales and registration data which incorporate all of the information rather than some subset of the information.

Another tempting fallacy is that there is a better forecast of demographic or economic factors than Urban Science has of automobile registrations or sales. A forecast of sales or registrations using sales or registration data alone is the best forecast of sales or registrations based upon available data. It is only the subjective inclusion of perceived future events that can appreciably change this forecast. Various methods may be used to incorporate this expert knowledge or opinion into the forecast. These methods include: structural model constraints, Bayesian priors and Bayes rule. If expert opinion is utilized in the methodology then it should be revealed to the client and if possible validated on historic data. One way to see the point above is to note that if you are using demographic or economic information for a forecast then that is a model forecast of the past and present data. It is understood in the banking industry that car sales in fact are the better predictor of economic conditions, rather than the other way around.

9. Model Verification

The automotive marketplace is a dynamic one. The manufacturers, products, sellers and buyers are constantly changing. As such, it is important that the model verification be as close in time as is relevant for a particular purpose. So for example, if one is forecasting out into the future then it is desirable that the model have internal and external consistency near the present point in time. In this case how the model performs on data that is 20 years old may not be irrelevant. On the other hand if one is concerned with determining the profitability of a marketing campaign of 20 years ago then of course the model fidelity in that 20 year old time period may be critical.

There is also the constant struggle between the opposing forces of as much recent data as possible for estimation versus the recent data being held out for verification.

10. Real World Solutions for Real World Problems

Now that we have categorized the real world problem by how it occurs within the auto industry, Figure 2, we next categorize the problem solutions and statistical models. The models utilize din the different areas may or may not be similar. The model used depends upon factors such as: available data, desired purpose, sophistication of the modeller and computing power. The following models are driven by solutions to the business aspects portrayed in Figure 2.

In order to create a statistical model the objective needs to be defined. This may be the estimation of the market share of a particular brand, the number of sales of a particular make model, the amount and time of a competitive incentive, the amount of inventory, etc. Once the objective is known then we need to do two things. The first need is to collect the value of the objective for existing data. The second is to collect information which can be manipulated to estimate the existing data objective using the information and statistical modeling techniques. The statistical modeling process is included within the solution creation process of Figure 4. The modeling process implicitly includes feedback.

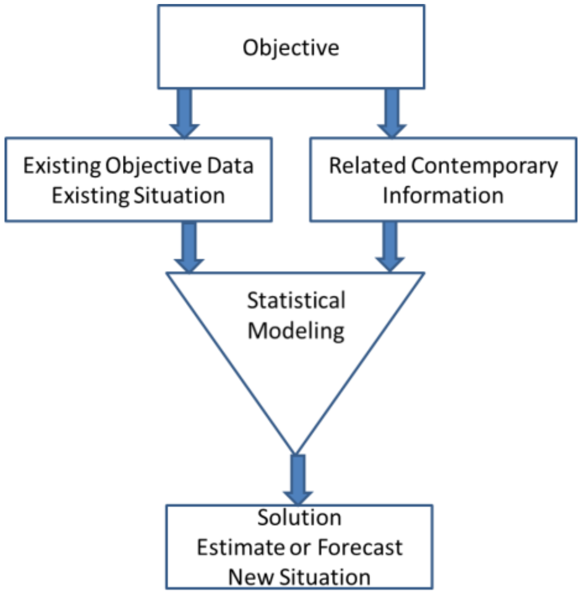

Figure 4: The Solution Creation Process (implicit feedback)

Objectives that occur in the automobile industry include: optimizing market share, sales and profitability. One issue with market share is that all vehicles are treated as 1 unit each. In fact some units may be worth more or less than other units. Even this has a temporal component. Is an inexpensive introductory small vehicle worth the same amount as an expensive luxury vehicle? In units the answer is yes. However, if one takes profitability and time into account then the answer may be different. With the inexpensive small car one may be acquiring a first time buyer and with loyalty may have them as lifetime buyers of multiple vehicles, possibly even culminating with the purchase of an expensive luxury vehicle 30 years into the future. With the expensive luxury vehicle this may in be their last purchase of a new vehicle, resulting in a nice short term profit but no longer term potential.

The standard way to handle this problem is to weight the vehicle purchases accordingly. This may involve the lifetime value of a customer or vehicle profitability. With profitability there are multiple components, for example, profitability to the seller and profitability to the manufacturer. These may be different things.

Once the objective has been determined then relevant data needs to be acquired and utilized. The relevant data take several forms. The most valuable is often the objective data. By objective is meant things that can be observed and counted. For example, number of vehicles sold, amount of money paid, etc. Sometimes objective data is not available. In this case surrogate data may be utilized instead. Often one prefers buyer behavior data over perceived behavior data.

11. The Entire System

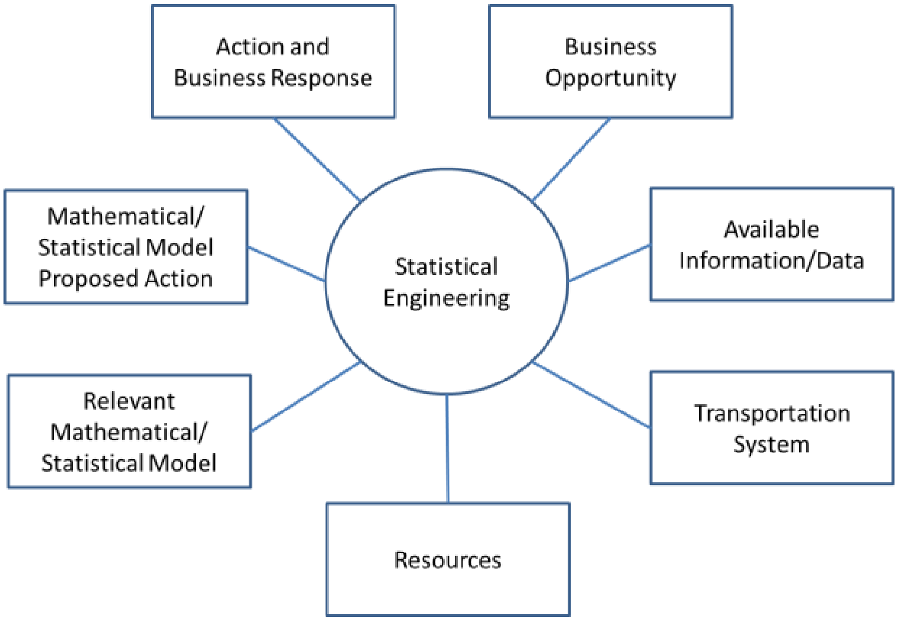

Figure 5: The Entire System

Figure 5 has a portrayal of how Statistical Engineering fits in and orchestrates the entire modeling/business opportunity satisfaction process. There is a "Business Opportunity" for a client. The opportunity and client may be internal or external to the statistical engineering organization. This opportunity can be capitalized on by providing the appropriate "Action and Business Response" to the client. This is accomplished by utilizing "Resources" to acquire the required "Available Information/Data", understand the real world "Transportation System", mimic that system with a "Relevant Model", utilize the model to determine a "Model Proposed Action" and then to provide the results of the modeling with a proposed real world "Action and Business Response".

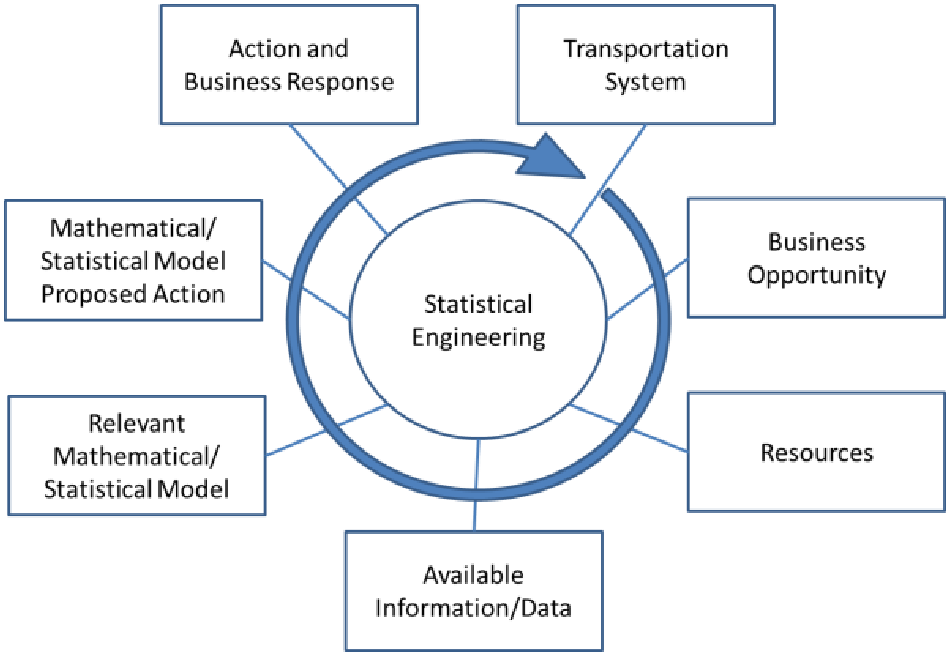

Figure 6: The General Circulation of the Entire System

The ultimate goal of this report is to understand which models are encompassed by the statistical engineering process and should be represented in the above "Entire System" of Figure 5.

12. Temporal Considerations

The data utilized in the experiment is either from the past present or future. The use of the model will be intended for past present or future. Both of these time periods are critical for system classification.

One of the most often cited forecasting issues is the desire to forecast registrations or new car sales into the future. To do this requires information about the future. This may be obtained by various methods. These methods include: expert opinion and or model assumptions. These methods are then combined with current data which is then somehow projected into the future. This current data may be the actual registrations or new car sales in question or other correlated or covariate data. In most cases it is the historical record of registrations or new car sales which contains the majority of relevant structure to do this forecasting properly.

In a few cases other information is relevant. These would include radical market or economic structure change, such as that caused by hurricane Katrina in New Orleans, hurricane Sandy on the upper east coast, the tsunami in Japan and a national significant recall which may have in fact stopped production or sales.

The client often asks for inclusion of external data that has little or no forecasting value above that of the historical registrations or new car sales. This data may be included in the modeling process. The estimation procedure is then used to select the appropriate weighting. The estimated weighting typically results in no weight to the covariate.

Sometimes one has external forecast values for the covariate. In a similar manner a model created from the historical automotive registrations or sales and the historical covariate data is often a better forecast of the covariate than the provided forecast. Often the estimated weighting will down weight the covariate as providing little value to the forecast accuracy of the covariate.
In addition one requires a historical forecast record to properly evaluate the use of the forecast values. This record is often unavailable for the time period required and even then it is often not available in the time granularity required to be of practical utility.
The bottom line here is that the historical record of automotive registrations or new car sales is one of the most valuable pieces of information or data for use in more accurately forecasting many attributes that are economically related.

13. Spatial Considerations

The importance of spatial analysis of the transportation market is legion. This is summarized in the phrase "location, location, location". In the automotive arena there are many locations that are important. The most important locations are those of the seller and the buyer. Seller locations include dealership, dealership satellite, buyer home, buyer work and buyer transportation locations.
The data coding these locations can be of many types and quality.

14. Spatial and Temporal Interactions

There are various possibilities of these interactions. These include dealer network changes over time, individual and aggregate buyer location changes. Sometimes these are considered constant in the time period and market considered and sometimes these are part of the modeling process. For example, analysis of a dealer network change may include the changes to the dealer network over time as well as buyer movement within the market over time.

15. Statistical and Mathematical Model and System Classification

It will be required to add the classification of utilized statistical and mathematical model systems. This classification will involve various statistical and mathematical models and various data cleansing and outlier detection methods.

16. Putting It All Together

With the classifications for the problems, information/data, solution systems we can put it all together in a statistical engineered system.

17. Agent Based Modeling

This document describes a proposed Engineered Model Vault and Dashboard. The model vault is meant to encompass all models utilized within the transportation statistically engineered system as one integrated all containing model. The model dashboard is meant to give access to each sub-model in a realistic data setting. By realistic data is meant to use real data if available and appropriate surrogate data or model fixes otherwise.

The model vault is at the opposite extreme of the data utilized for analysis. It is contains all models desired for enterprise model solutions.

One example of a model vault type model is the Urban Science GainSmarts which includes many sub-models for the data mining prediction and then checks this prediction against a holdout sample.

The model vault will have as its focus the location aspect of our business. Upon this base is built all of the other possible models that are considered for use for in the enterprise solutions. With this model structure in place the enterprise will accrue many benefits. These include:

1. A census of possibly valid modeling techniques
2. The interdependence of solutions on relevant factors becomes obvious
3. A dashboard type visualization of client and competitive actions and responses
4. Model building knowledge is retained in one location and for all time
5. Simulation market model

A prototype of this modeling vault and particularly dashboard idea has been constructed in the Massachusetts Institute of Technology software called StarLogo. This software is an agent based software. Being primarily an education tool the prototype stretches the limits of this software and yet, with a little imagination, demonstrates the possibilities available for the model vault.

In order to develop this idea to a more beneficial method will require construction in a more sophisticated agent based software. Possibilities for this software include: Mason, Netlogo, Repast and Storm. Further information may be found at: http://www.grids.ac.uk/Complex/ABMS/.

This type of analysis is becoming more popular with the increasing capabilities of computer hardware and software. The competitors of the enterprise will soon be using this modeling methodology. By using it ourselves we will prevent unwanted loss of business.

The current prototype requires the implementation of the MIT software StarLogo. The prototype model vault and dashboard includes:

1. Dealership attractiveness modeled with a distance decay function and the resulting effect on sales.
2. Three dealerships in the market
3. Locations of buyers in the market
4. Various geography definitions
5. Introduction of Advertising amount by dealership
6. Introduction of Inventory by dealership

7. Daily time progression of sales based on a sales propensity model
8. Ability to assign distance and decay attractiveness by dealership
9. Output includes: Sales by dealership/brand by day
10. Output includes Market Share by dealership /brand by day
11. Display profitability of the dealership based upon the chosen specific parameters (attractiveness, decay, inventory, and advertising).

The prototype may:

1. Reflect outcome of sales/market share/profitability by day
2. Be used in a simulation model showing random variability on the outcome based on specific parameters
3. Be used in a client action effect on the outcome
4. Be used to determine competitive actions on the outcome

18. Summary

The humans involved in this automotive marketplace process may are analogous to the atoms and the consumer behavior is analogous to the physical forces referred to in Stringer (2008). Continuing with this analogy practical use of this modeling methodology requires vast computing resources as described in Stringer (2008).

Acknowledgements

This work was accomplished while at Urban Science Applications, Inc.

References

Lampathaki, Fenareti, Sotiris Koussouris, Carlos Agostinho, Ricardo Jardim-Goncalves, Yannis Charalabidis and John Psarras, 2012, "Infusing scientific foundations into Enterprise Interoperability", Computers in Industry, 63, 858–866.

Edmund, Norman W. (2012): SM-14 http://scientificmethod.com/b_index.html

Stringer, Steve (2008), "Connecting Business Needs with Basic Science", Research-Technology Management, Vol. 51, No. 1.