Interactive Visualization and Statistical Interpretation of Market Share, Competitive Price, Stock, and Mortgage Loan Data Using JMP® 3.1

James Daughtry, Principia Information Research Group
Dr. Jonathan Pinder, Wake Forest University
Dr. Edwin Tang, Chinese University of Hong Kong
William Gjertsen, SAS Institute Inc.

Introduction and Abstract
The Information Age is here! The "Information Highway" and all of its features including market data retrieval, the downloading of text and images, real time payment via credit/debit cards and just in time production and inventory control are part of our current business world. However, whatever the source, we will still be presented with business and market data that needs to be visualized, analyzed, and understood before it becomes meaningful information.

New interactive computer tools, such as JMP, make this visualization and understanding possible. This presentation will draw on dynamic linking of screens, the 1, 2, 3 and N-dimensional representations of outliers and the analytic discovery tools available in JMP. Included will be examples for 1) Market Share and Competitive Price Analyses, 2) Stock Tracking and Selection and 3) Mortgage Loan Portfolio Analysis using Multinomial logit analysis. Consideration will be given to how the data can be retrieved, structured, and transformed. Data access and retrieval into JMP relational data tables will be discussed.

Market Share and Competitive Price Analyses
All business enterprise in the USA and the "Global Market" is currently deluged with data. The bombardment of data consists of thousands and millions of records. These records are for instance: financial, accounting, customer, inventory, purchasing, production, manufacturing, market, and sales in nature. The data sources can be internal and external; national and international. For the government the source can be city, county, state and federal. The mode of data can be cross sectional or longitudinal in time, and can appear as text, numbers, or multimedia images.

A corporation's top management: the board of directors, the chief executive, financial, and information officers, all need high technology solutions that can analyze data easily and present it as information. JMP is an excellent analytical engine to implement the conversion of data to information.

To the CEO, the consumer and the investor community, market share is one of the most important "bottom line" financial measures of a successful business. Witness the recent concern of Mr. Michael Spindler, CEO, Apple Computer Company, when Compaq knocked Apple out of first place in market share of the PC market. His response was to set new goals and rejuvenate Apple's sales force in an effort to regain market share over the next few years.

For any market share analysis it is first important to define the market. In our example the market definition is the total resurfacing and construction program for the highway system of an entire southern state. The data is first entered and edited as SAS data sets and then brought over to JMP for data exploration and outlier detection.

The data originated as SAS data sets but was outputted to us in a PC DOS formatted file as a text file with SAS PUT statements. The PC DOS diskette was read directly into Microsoft Word® on a Macintosh with a Power Macintosh 7100/66. Word let us view the incoming text file and confirmed that backslash, \ , was the field delimiter. Next it was easy to import the data directly via the Text selection under Import within JMP. We even kept the same variable names by choosing Label within the import dialog. The only modification occurred with amounts of money which were in a SAS Dollar format (leading $ sign and commas). We had to read a number such as $60,448.50 as a character
variable and create a second variable which is numeric that eliminates the $ and , entries and converts it to a continuous number such as 60448.50. This is done with the JMP Calculator formula as shown:

\[
t1_{-} \cdot 0 \\
\sum_{j=2}^{4} t2_{=} \text{num item}(j, \text{of} \ BTEXTAMT, \text{delimited by "",}) \\
t1_{=} \begin{cases} 
  \{t1 \cdot 1000 + t2, & \text{if } t2 \neq 0 \\
  t1, & \text{otherwise}
\end{cases}
\]

results t1

The JMP Tables retrieved consisted of the following files:

- DBIDTABS (and DBIDTABS2): Bid tabulations on all items.
- DPRJTQ: Project items. Item quantities and prices of all projects that were let.
- ITEMLST: A list of construction items per contract.
- VENDOR: A list of vendor IDs and names.
- VENDADD: A list of vendor addresses.

The market share analysis provides comparative market share for all contractors, subsets of contractors, over different time periods and different counties within the state. The market share data is expressed as: number of contracts bid, number of contracts awarded, total market dollars, percentage of dollars per contractor, total dollars and percent of dollars for particular construction items by contractor. Market share analysis over periods of time for all competing contractors is very important to all bidders. Analysis over time indicates the stability of a contractor's market share and determines its present and future capacity. Market share is highlighted in Figures 1, 2, 3.

Another important issue for corporate CEOs is competitive price analysis. This is particularly true in our case where contractors must bid on complete construction jobs. To launch a new product or be the low bidder in the marketplace it is imperative for success to have a firm understanding of the competitive price at the item level. Item price analysis can include average summary statistics, regression analysis, cluster analysis, and control charts of competitive bids and awards as in Figure 4. In many cases price-quantity relationships can be analyzed in a cross section of time, over a range of time, in a geographic region, for a subset of competitors or the whole market, and for a particular item or group of items. Figure 5 is an example of portraying the price-quantity relationship for asphalt cement (AC) bids for the year of 1992. As expected as quantity increases price per unit should decrease and interesting competitive prices and patterns can be identified.
Figure 1: Vendors with highest market share of individual contract awards, a first look at the data.
Market Share Leaders by Award Amount and Number of Contracts

Figure 2: Display the Market Share Leaders and select the Top 10 by drawing a reference line at $5,000,000 and label each company with # of contracts.

Top 10 in Total Dollar Market Share

Figure 3: Our Top 10 Vendors captured 52.5% of total $146.69 million market.
Figure 4: Bid Unit Price vs. Contact ID "Control Chart"

Figure 5: Unit Price vs. Quantity for Asphalt Cement Contracts for 1992
Stock Tracking and Selection

An application well suited for JMP is market timing, technical analysis of stocks and selection of a stock based on your own fundamental criteria.

"The stock market is an incredibly inefficient engine for making money. Market timing is merely an attempt to try to make it more efficient. ¹ Whether you are a portfolio manager or an individual investor its very important to know not only the individual equity's characteristics but the profiles of the sector and overall market as well. You don't want to invest in a stock, no matter how attractive, in an overall bear market.

We will first illustrate technical analysis indicators that are easy to implement within JMP for the overall market and apply similar ideas to track an individual stock. We will first look at a few analytic and graphical tools for momentum and overbought/oversold conditions and then apply some of these same ideas to a particular stock.

Market timing using trend line analysis:

The Cumulative Haluran Index (Cum HST) has a high correlation with DJI.

Given that this indicator is highly correlated with the Dow Jones Industrials index, JD, we can use it to illustrate trend line confirmations and divergences, our first market timing technique.
Drawing correct trend lines is one of the easiest and most powerful tools available to a technical analyst. The way to draw correct trend lines is described by Speared:

"For an uptrend within the period of consideration, draw a line from the lowest low, up and to the highest minor low point preceding the highest high so that the line \textit{does not pass through prices in between the two low points}. For a downtrend, draw a line from the highest high point to the lowest minor high proceeding the lowest low so that the line \textit{does not pass through prices in between the two highs}.”

The above two charts indicate a change of trend from down trend to uptrend since the correct short term trend lines have been broken through in an upward direction. Use the Annotate or “A” tool in JMP to draw correct trend lines.

\textbf{Polarized Fractal Efficiency}

Although trend lines can and should also be used effectively with individual stocks, it is instructive to introduce another tool, an oscillator based on fractal dimension. A fractal is an object in which the parts are in some way related to the whole. Trees branch according to a fractal scale. Each branch, with its smaller branches, is similar to the whole tree. The time series of a stock or index in the same way has self-similarity with respect to time. Below is snapshot of daily and weekly charts from the same stock. Can you tell which is which?
The Polarized Fractal Efficiency measures price change distances two ways and takes the ratio to create an oscillator. When the oscillator crosses above 0 from below it is a buy and when it descends below 0 from above it is a sell. The FE numerator measures the “long ruler” straight line distance, $F$, between Close$_i$ and Close$_{i+10}$ and each $D_j$ distance in the denominator is the “short ruler” distance between Close$_i$ and Close$_{i-1}$. The FE denominator, $G$, is just the overall “short ruler” distance between Close$_i$ and Close$_{i+10}$. FE is then smoothed with a 5 period EMA. The smoothed FFE will oscillate between $+100\%$ and $-100\%$.

\[
F = \sqrt{(CSCO_i - CSCO_{i-10})^2 + 100}, \quad \text{if } i \leq 10
\]

\[
G = \sum_{j=(i-9)}^{i} \sqrt{(CSCO_j - CSCO_{j-1})^2 + 100}, \quad \text{otherwise}
\]

\[
FE = \begin{cases} 
F \cdot 100, & \text{if } CSCO_i \geq CSCO_{i-10} \\
-(F/G \cdot 100), & \text{otherwise}
\end{cases}
\]

\[
PFE = \begin{cases} 
(FE_i \cdot 0.333 + PFE_{i-1} \cdot 0.667), & \text{otherwise}
\end{cases}
\]
Here we see the benefit of market timing vs. just buy and hold. If you are aggressive, make each sell a sell short signal as well as sell the stock you bought previously signal. So the PFE market timer sells the stock at 34.75 that he bought at 27.875 and sells it short at 34.75 and buys it back later in 1994 at 24.625. The net gain is $6.875 + 10.125 = 17$ for the market timer and $25.675 - 24.625 = 1.0$ for the buy and hold individual. Note that basic trend line analysis should also be used on Cisco prices. Weinstein offers yet another set of buy/sell rules based on stage analysis which can act as a second confirmatory tool to PFE rules. In practice the technical trader uses several indicators (i.e. trend lines and oscillators) simultaneously to make market decisions.
A stock selection example:
From our universe of stocks to be selected (i.e. Investor's Business Daily highlights a new industry group every day), choose those that have a percentile ranking of 80 or more for both earnings per share (EPS Rnk) and relative strength to all other stocks (Rel Str). We can use the dynamic linking between histograms to visually make this first filter and subset the data. Then spin the resulting dataset using fundamental indicators and discover hidden gems like Micron Technology! Then begin tracking that stock in earnest.

**EPS Rnk**

**Rel Str**

**Components**
- Net % Prof Margin
- Ret on Equity
- Undervalued index
- Last Q Sales

**Mortgage Loan Portfolio Analysis**

*Multinomial Logit Model*

In a situation described involving loan accounts, there are three states, rather than two, in which an account may be categorized: default, prepayment, or an active. To extend the two-state model to a set of three alternatives, with no particular ordering, let: $Y_{ij} = 1$ if the $i$th observation chooses alternative $j$, $j = 1,2,3$. In the situation of interest here, $i$ represents the accounts in a portfolio and $j$ represents the states of default, prepayment, or an active account. In a similar fashion as above, let $p_{ij} = P(Y_{ij} = 1)$, then for the three-state model:

$$p_{11} + p_{12} + p_{13} = 1,$$

which generalizes to

$$\sum_{j=1}^{J} \pi_{ij} = 1$$

for the general case with $J$ alternative states for the qualitative
dependent variable. Thus, for the generalized multinomial model:

\[ \pi_{ij} = \frac{\exp(\alpha_j + \beta_j x_i)}{\sum_{j=1}^{l} \exp(\alpha_j + \beta_j x_i)} \]

In the three-state model, solving for \((\alpha_2 + \beta_2 x_i)\) and \((\alpha_3 + \beta_3 x_i)\) yields:

\[ \log \left( \frac{\pi_2}{\pi_1} \right) = \alpha_2 + \beta_2 x_i \]
\[ \log \left( \frac{\pi_3}{\pi_1} \right) = \alpha_3 + \beta_3 x_i \]

\[ \pi_1 = 1 - \pi_2 - \pi_3 \]

where \(\alpha_1 = \beta_1 = 0\) and state one is the benchmark for comparison (Theil (1969)).

To demonstrate the implementation of the decision analysis/multinomial logit model, consider the following specific loan portfolio circumstance. The loans in this portfolio are for mobile homes and involve three parties: the mobile home purchaser ("the customer"), the mobile home seller ("the dealer"), and the loan originating institution ("the bank"). State lending regulations prevented banks from granting loan terms longer than 12 years for mobile homes. Therefore, the bank created a "dealer participation plan" in which the customer received a 15 year loan, of which the bank received the first 12 (satisfying the regulations) and the dealer received the final 3 years of payments in exchange for acting as guarantor for the customer. Thus, in the event of default during the first 12 years, the dealer was responsible for repaying the balance of the principal. In the event of prepayment, the dealer received the difference between the customer's principal balance and a lesser balance for which the dealer was guarantor.

For several of reasons dealers failed to honor their financial obligations of the participation agreement. At which time, in accordance with the contractual agreement, the bank seized the rights to the 'participation' account payments in a failing dealer's portfolio. These payments then became the subject of legal proceedings. Thus, a valuation was necessary to determine a fair market value of portfolios for which there is no existing market.

In one failing dealer's portfolio there were a total of 89 participation accounts. Of these accounts, 28 had either prepaid or defaulted prior to the dealer's failure and comprised a set of closed accounts. The dealer filed a claim of $628,285 for the 3 years of payments from the remaining 61 open accounts. Thus, a valuation was prepared for the 61 open accounts and was dependent upon estimates of the effects of future defaults and prepayments.

To begin the valuation process, the two sets of accounts were combined and the status of each account was categorized as either default, prepaid, or open. In order to estimate the multinomial logit model, variables had to be available for each type of account. The variables available from the bank's database were:

1. **Status** - Account status: default, prepaid, or open; the dependent variable.
2. **Date** - Loan origination date.
3. **Balance** - Principle balance owed by the customer.
4. **RemPmts** - Number of remaining payments.
5. **IniRate1** - Original interest rate.
6. **IniRate2** - Current interest rate (for accounts with variable interest rates).
7. **30** - Number of payments 15-30 days late.
8. **60** - Number of payments 31-60 days late.
9. **90** - Number of payments 61-90 days late.
10. **90+** - Number of payments more than 90 days late.

To develop the multivariate multinomial logit model, each of variables were test in various models and the significant variables reduced to a model containing the independent variables 60, RemPmts, and IniRate2. The results of this model are given in Table 1.
To determine the state prediction (\(p_{ij}\)) equations, let the states default, prepaid, and open be represented by the indices 1, 2, and 3 respectively and let state 1 (default) be the benchmark state (parameters = 0). From the parameters in Table 1 are used to calculate the linear terms:

\[\text{Linear}_2 = 28.956 - 1.9755 \times 60 - 0.13792 \times \text{RemPmts} - 219.75 \times \text{IntRate2}.\]
\[\text{Linear}_3 = 49.115 - 1.9755 \times 60 - 0.13792 \times \text{RemPmts} - 219.75 \times \text{IntRate2}.\]

Using these linear terms the equation for estimating the \(p_{ij}\)'s are:

\[p_{i1} = \frac{1}{1 + \exp(\text{Linear}_2) + \exp(\text{Linear}_3)}.\]
\[p_{i2} = \exp(\text{Linear}_2) p_{i1}.\]
\[p_{i3} = \exp(\text{Linear}_3) p_{i1}.\]

Rather than using these estimates of \(p_{ij}\) to forecast of the specific state of each account, the \(p_{ij}\)'s measure the state probabilities associated with each account. Thus, they quantify the potential that every account has for defaulting, prepayment, or remaining open.

After building the multivariate multinomial logit model, univariate multinomial logit models were estimated to illustrate the qualitative effect on the status of each account of each significant independent variable in the multivariate model. The univariate logit models for are illustrated in Figures 6-8. The vertical spacing between the logit curves represents the likelihood of each state and is measured on the left-hand scale. The right-hand scale indicates the overall proportions of the three states. Thus, if the dependent variable were not associated with the independent variable then the logit curves would be horizontal lines located at the hash marks separating the three states.

Figure 6 demonstrates that the likelihood of default becomes critical at about 4 payments 60 days late. Furthermore, at about 9 payments 60 days late the likelihood of default is nearly certain due to the collection practices of the bank. Figure 7 shows that as the number of remaining payments decrease there is less likelihood of default and that the likelihood of prepayment remains constant over a broad range of remaining payments. Figure 8 shows that as the current interest rate increases there are sharp increases in the likelihood of prepayment and default and that approximately 14% is the critical rate at which customers shift towards prepayment and default.
After estimating the $p_{ij}$'s, the next step in the valuation process was to estimate the net present values (NPV$_{ij}$) of each state for each account. As before, let the states default, prepaid, and open be represented by the indices 1, 2, and 3 respectively. For each open account, NPV$_{11}$ was determined by subtracting the principal balance and the estimated cost of repossession from the estimated proceeds of selling the repossessed asset. The cost of repossession was estimated from the costs of the previous defaults. Based on the selling price of the previous defaults, at the age of the asset, a logarithmic model was regressed to estimate the repossession sales proceeds.

NPV$_{12}$ was calculated as the difference between the customer's principal balance and a lesser balance for which the dealer was guarantor. Finally, NPV$_{13}$ was calculated as the present value of the final 3 years of payments to be received by the dealer if the loan proceeded to full term.

Thus, the portfolio valuation for the 61 open accounts was calculated by first estimating the state probabilities, $p_{ij}$, for each account. Next, the net present values (NPV$_{ij}$) for each state for each account are computed as described above. Subsequently, the expected NPV for each account, E[NPV$_{ij}$], is determined by:

$$E[NPV_{ij}] = \sum_{j=1}^{3} p_{ij} \cdot NPV_{ij}$$
Finally, the portfolio valuation is calculated by summing $E[NPV_i]$ for all accounts. Thus, the portfolio valuation is based on the relative risks inherent in each account rather than trying to specifically predict which accounts will default or prepay. Using this method for the set of accounts in this example, the portfolio valuation was calculated as $267,497; well below the $628,285 claimed by the dealer in the lawsuit.

References for Stock Timing and Selection


Multinomial Logit Reference