**EXPLORATORY VISUALIZATION OF MULTIVARIATE DATA: THE CASE OF NIKKEI INDEX**

*Abbas Heiat, College of Business, Montana State University-Billings, 1500 University Drive, Billings ,MT 59101, 406-657-1627, aheiat@msubillings.edu*

**ABSTRACT**

Discovering factors that influence the stock prices and relationships among them is difficult and complicated. Many researchers believe that representing data visually should improve business analysis and decision-making. In this paper, we apply Scatter Plot Matrices, Parallel Coordinates, Glyphs, and Dimensional Stacking to visually explore a multivariate stock market and financial time series data set.

**INTRODUCTION**

Many investors, including individuals and institutions, are interested in forecasting stock prices in the financial markets. Discovering factors that influence the stock prices and relationships among them, however, is difficult and complicated. In the business field, the commonly used visualization methods by analysts and decision makers are two dimensional line, bar, and pie charts. These tools are used in combination of statistical modeling and estimation methods to analyze important factors and make business decisions. These techniques, while well known and easy to use and understand, are not adequate for discovery of patterns and relationships in multivariate data sets. In recent years, there has been a rapid growth in the field of scientific multivariate visualization. New powerful multi-dimensional tools are available in the marketplace. But their applications in the business field have been slow and limited. In this paper, we apply Scatter Plot Matrices, Parallel Coordinates, Glyphs, and Dimensional Stacking.

# **DATA**

# The data is collected via the Internet. Data for variables were collected for a week that was selected randomly. A total number of 262 observations were included in the data set. The study used one dependent variable (the Nikkei Average) and five independent variables as follows:

1. The Exchange Rate between Japanese Yen and the U.S. Dollars
2. The Federal Fund Rate in the U.S
3. Dow Jones Industrial Index
4. NASDAQ Index
5. SP500 Index

The exchange rate represents volume notation, which is the Japanese Yen over the U.S. dollars. In other words, the value represents the value of one U.S. dollar in terms of the Japanese Yen. Federal fund rate is the interest rate determined by the Federal Reserve for a short-term (usually one day) lending of federal funds among financial institutions (banks).

**MULTIVARIATE VISUALIZATION TECHNIQUES**

In this paper we have used XmdvTool developed by Ward et al (1996) for multivariate data visualization. XmdvTool is a public-domain software package for interactive visual exploration of multivariate data sets. Four distinct methods of multidimensional projections are provided: Scatterplots, Glyphs, Parallel Coordinates, and Dimensional Stacking. Users can switch between display techniques by simply clicking on button in the interface. A major tool in XmdvTool for providing insights into N-Dimensional spatial relations is the N-Dimensional brush, which allows users to perform operations on the data points which fall within a user-specified N-Dimensional subspace of the total space defined by the data (XmdvTool Help, 2008). In this paper we are going to use the first three methods for representing our financial data set.

**Parallel Coordinates**

In Parallel Coordinates, each dimension (attribute) corresponds to an axis, and the N axes are organized as uniformly spaced vertical lines. A data record in N-dimensional space manifests itself as a connected set of points, one on each axis. A point in Cartesian coordinates corresponds to a polyline (connected set of points) in Parallel Coordinates. Points lying on a common line or plane create readily perceived structures in the image. In generating the display of parallel coordinates in XmdvTool, the view area is divided into N vertical slices of equal width. At the center of each slice an axis is drawn, along with a label at the top end. Data points are generated as polylines across the N axes. Figure 1 and Figure 2 illustrate the use of Parallel Coordinates to help to visualize the relationships among variables in our data set. In the Hierarchical Parallel Coordinates the clusters rather than individual data records are displayed. The mean of a cluster is mapped to a polyline traversing across all the axes(Ward, 2008).

**Figure1. Parallel Coordinates for representing relationship among market financial data.**

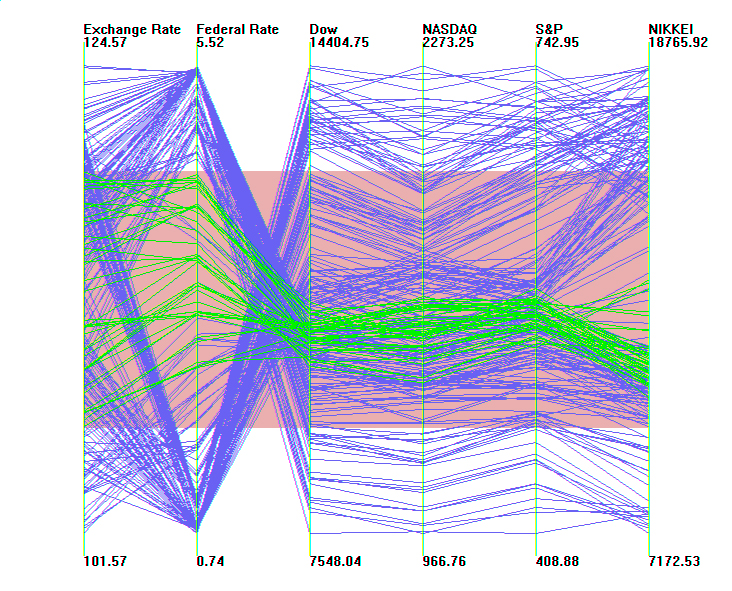
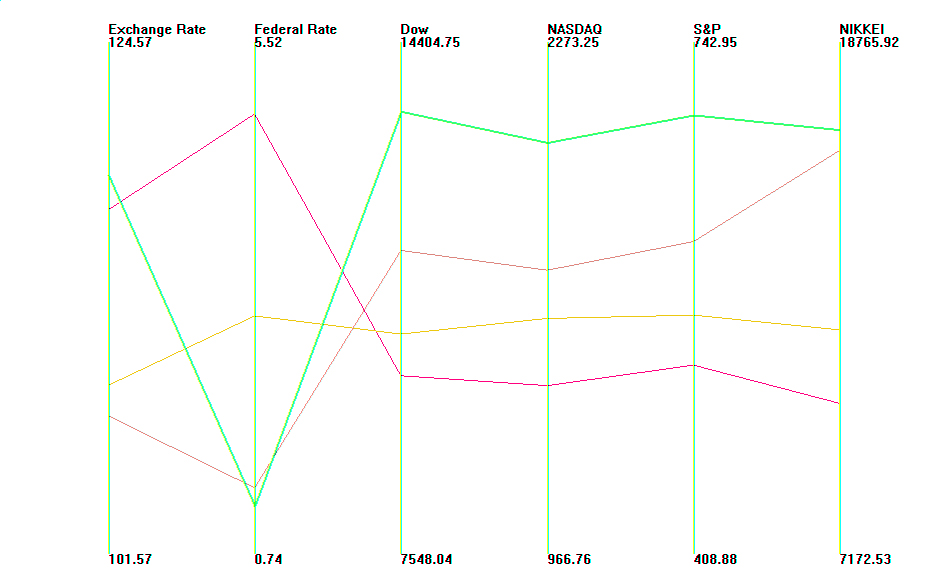


Figure 1 and Figure 2 show, there is some positive correlations between exchange rate and federal rate. There are strong negative correlations between exchange rate and federal rate and the rest of the variables (stock market prices). Figures also indicate strong positive correlation among different markets.

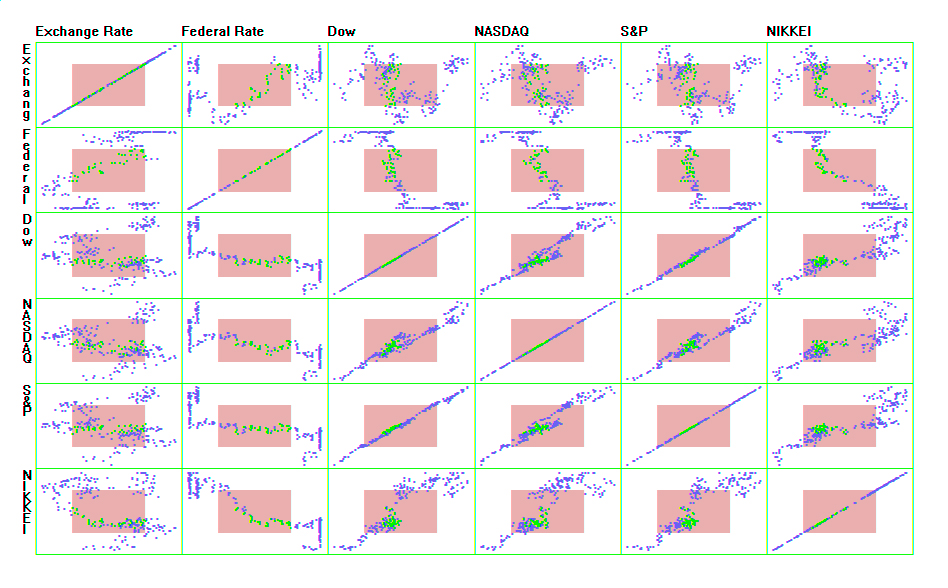
**Scatterplots**

Scatterplots are one of the most commonly used methods to project high-dimensional data to 2 dimensions. In this method, a grid of parallel projections of the data is generated. Each of these is a simple plot for the two dimensions it represents. The horizontal dimension of a scatterplot is controlled by the column it resides in, and likewise the vertical dimension is controlled by the row. At the top of each column and at the left of each row is a label that shows what dimensions that row / column represents. The basic idea is to visually compare features in one panel with features in others. The patterns or trends can be detected in both X and Y dimensions. Figure 3 not only confirms the observations made about the financial data set through Parallel Coordinates, but it makes it possible to make an additional important observation. The negative correlations that exist between exchange rate and

**Figure2. Hierarchical Parallel Coordinates for representing relationship among market financial data.**

****

**Figure 3. Scatterplots for representing relationship among market financial data.**

****

federal rate and different market prices are not continuous. In other words, markets do not respond to small rate cuts and exchange rate fluctuations until they become significant changes (Ward, 2008).

**CONCLUSION**

In this paper, we used three techniques to visually represent and explore relationships among six financial market variables. These visual representations helped us to gain insight about patterns in our data set. We identified four clusters and nonlinear negative and positive correlations among the variables. Specifically, visualization suggested that markets do not respond to small rate cuts and exchange rate fluctuations until they become significant changes.