Abstract
Visual data analysis has received a lot of research interest in recent years, and a wide variety of new visualization techniques and applications have been developed to improve insight into the various application domains. In financial data analysis, however, analysts still primarily rely on a set of statistical performance parameters in combination with traditional line charts in order to evaluate assets and to make decisions, and information visualization is only very slowly entering this important domain. In this paper, we analyze some of the standard statistical measures for technical financial data analysis and demonstrate cases where they produce insufficient and misleading results that do not reflect the real performance of an asset. We propose a technique for visualizing financial time series data that eliminates these inadequacies, offering a complete view on the real performance of an asset. The technique is enhanced by relevance and weighting functions according to the users’ preferences in order to emphasize specific regions of interest. Based on these principles we redefine some of the standard performance measures. We apply our technique on real world financial data sets and combine it with higher-level financial analysis techniques such as performance/risk analysis, dominance evaluation, and efficiency curves in order to show how traditional techniques from economics can be improved by modern visual data analysis techniques.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation H.4 [Information Systems]: H.4 Information Interfaces and Presentation

1. Introduction
Many different sophisticated visualization methods have been proposed in the past to analyze time related data sets in various application fields [SM03]. In the area of financial analysis, the most important and still almost exclusively used visualization techniques are line charts, bar charts, point charts or variations hereof, which are usually used in combination with statistical performance measures, such as 1-year, 3-year or 5-year performance, or the volatility of an asset. The techniques are utilized by financial analysts to evaluate whether a bond, stock, future, commodity or market index is worth an investment. In this paper, we show that the statistical measures may produce misleading results that can lead to wrong conclusions and propose a new visualization technique for financial analysis that will support the analysts to overcome these problems. Our visual analysis technique offers better insights into the behavior of an asset than numeric performance measures and also supports user definable relevance weighting to emphasize specific regions of interests. We integrate our technique with well-known financial analysis techniques, such as performance/risk analysis, dominance plots, and efficiency curves. Our approach provides a powerful method for analyzing financial time series data, allowing a more complete analysis than with existing statistical performance measures or previous visualization approaches. We present real world market application examples and demonstrate the potential of our approach for financial data analysis.

2. Background
2.1. Analysis of Financial Time Series Data
In the area of financial data analysis, the visualization techniques to analyze assets have not changed much and almost remained the same for decades. Newspapers and financial magazines still primarily rely on line charts, bar charts or point charts [Mur99]. These techniques, commonly in form
of line charts, have the advantage that they are well known, easy to interpret, and they provide a simple graphical representation that can be intuitively used. Examples of these types of charts can be found on many financial web-sites, like OnVista [OV], MorningStar [MS], or any other online brokerage or financial portal on the web.

The simple visualization techniques are usually enriched by a set of statistical parameters, which typically include the 1-year, 3-year and 5-year performance, risk (volatility), maximum loss, average loss/gain during recession/booming period, and others. In the area of financial data analysis, combinations of charts and statistical values are widely used for the decision making process whether to buy or sell an asset. Especially the 1-year, 3-year and 5-year performance measures are often used to determine whether an asset is a promising and stable long term investment. This also illustrates a general difference from other time series data analysis fields, as the financial sector is particularly interested in pairwise combinations of two points of a time series (entry time and holding time).

2.2. Inadequacies of Financial Standard Measures

The statistical measures for financial data analysis have several problems and inadequacies. The yearly performance of an asset for example is simply determined by the growth or decline of an asset within a specified time interval. Considering the 1-year, 3-year and 5-year performance of an asset for example, these statistical values only reflect the performance of three very specific time periods, which are only measured relative to the current date. Between these time periods, the performance of an asset could vary significantly without having any impact on the performance measure. We illustrate this phenomenon in Figure 1 with a simple example of a technology fund during the ’dot-com crisis’, with the three time periods depicted above the chart and April 2002 on the right border as the reference date for the time of sale. By visually analyzing the price chart in Figure 1, it can easily be observed that this technology fund offered the opportunity to make a lot of profit if the fund had been purchased before 1999 and sold in the middle of 2000, as well as a high risk of losing most of the investment if the fund had been purchased in 2000 on the high peak and sold afterwards. However, in the example shown in Figure 1, neither the 1-year nor the 3-year nor the 5-year holding periods reflect the high potential of the fund in the year 2000. If a second example is taken, and September 2001 happens to be the reference date, the performance values would be completely different (\(\text{perf}_1 = -58\%\), \(\text{perf}_3 = +21\%\) and \(\text{perf}_5 = +50\%\)). In addition, minimal changes in the holding period or a slightly different reference date may cause the statistical parameters of volatile assets to change significantly by coincidentally hitting local peaks or lows. An example for this is the ‘Black Monday’ on October 19th, 1987, where the stock market lost 25% of its value in just one day, and the performance measures therefore return completely different results if the holding period or reference date is modified by just one single day.

These observations show that the three most widely used performance measures (1, 3 and 5 year performance) in financial analysis represent just three single points from a large set of performance measure, and may provide unstable and unreliable information about the characteristics of an asset which makes them inappropriate for making important investment decisions.

![DWS Technology Fund](image)

Figure 1: Price chart illustrating a technology fund during the ’dot-com’ crisis. The common statistical measures for 5-year, 3-year and 1-year performance (see table) fail to recognize the peak in the year 2000 because the peak is completely situated between the corresponding time intervals. If the reference point for the time of sale is shifted for only a short period of time (see red colored time intervals), the performance measures return completely different values although most of the chart is identical.

2.3. Related work

The financial sector is an important domain dealing with complex time dependent data sets, and in recent years several different approaches for visualizing these data have been developed. 'Financial Viewpoints' [SL95] for example allows a 3-dimensional flight through financial spreadsheets, other approaches are based on Self-Organizing Maps [BIS’95] [SK03] [DK98] or recursive patterns [KAK95]. One early approach based on treemaps [WM99] became popular after being commercialized in form of the ‘Map of the Market’ by SmartMoney [SM], where each rectangle represents an asset and rectangles are organized by industrial market sectors, with the size of a rectangle representing the market capitalization and the color reflecting the loss or gain, giving a complete overview of the financial market in one single view. Neovisions RiskMaps [NV], based on...
HeatMaps which organizes data in real-time by means of a regular matrix of colored rectangles, change the color and intensity of the cells in real-time to indicate shifts in the value of an asset. Other commercial products or web-based applications are GSphere by Gravity Investments [GRV], Market Topology [MT], Portfolio Impact by High Tower Software [HTS] or StockVis [PE] which presents multiple time periods in form of web-enabled 3D-charting. TimeSearcher [H03] provides an interactive tool for exploring time-series data by using timeboxes with techniques like details-on-demand. ThemeRiver [HHN99] has also been applied on financial time series data to identify increases or decreases of asset prices. One of the most recent approaches is based on triangular Growth Matrices which allow a comparison of one asset against a database of 12,000 assets and show the relative performance of one asset against the market for all time intervals in a single view [KNS06]. Except the last, none of the approaches mentioned above allows a visualization of a wide range of time periods.

3. Visualization of Financial Performance Measures

3.1. Conceptual Model of our Approach

To overcome the inadequacies of the financial standard measures explained above, we propose a visual approach that allows a greatly improved analysis and exploration of the long-term behavior of assets.

Our visualization is based on a pixel-based paradigm which consists of a two-dimensional rectangular box model in a Cartesian coordinate system with orthogonal axes, where the y-axis reflects all possible holding periods of an investment from 1 month to 5 years in monthly increments, and the x-axis reflects all possible reference dates of an asset (time of sale) from the current date to the past which at the same time serves as the reference date for calculating the time of purchase. Figure 2 illustrates the concept of our model, the point a, for example, denotes the 3-year performance of an asset with a time of sale in January 2000, and a time of purchase in January 1997.

Therefore, instead of covering a set of only three measures for three specific holding periods, our approach covers a large range of holding times (60 holding times by default) on the y-axis and more than 8 years as times of sale on the x-axis (100 months by default). If the default values are used, our visualization shows a total of 6000 performance values, allowing a simultaneous comparison of many investment horizons and a much improved analysis of a long-term behavior of an asset in just one view. The previously explained 1-year, 3-year and 5-year performance values represent one pixel each in our Performance Matrix. In order to illustrate this, the three measures PM(2005, 1y), PM(2005, 3y) and PM(2005, 5y) are denoted in Figure 2.

The performance for each combination of holding period and time of sale can be described as a metric function

\[ PM^A(h, s) = \frac{VA(s) - VA(s-h)}{VA(s-h)} \quad s, s-h \in T \]

where \( s \) is the time of sale and \( s-h \) is the time of purchase of an asset \( A \) taken from the set of all possible points of time \( T \), with \( h \) representing the holding period. \( VA(s) \) and \( VA(s-h) \) are the observed prices of asset \( A \) at times \( s \) and \( s-h \). The performance value is then annualized by

\[ PM^A_{\text{annual}}(s, h) = PM^A(s, h)^\frac{12}{T} \]

in order to calculate the effective performance per year, because profits in the first year of the holding period could be reinvested for making profits in the second year. For a holding period of \( h = 12 \) month, the exponent turns 1.

Figure 3: Our colormap visualizes the performance of an asset with a linear spectrum from red (loss) to yellow (neutral) to green (profit), capped at -60% and +60%.

Each performance value \( PM^A_{\text{annual}}(s, h) \) at a coordinate \( (s, h) \) is visualized by a colored pixel in our box model. Our colormap (see Figure 3) representing the performance values is a smooth transition from red (losses) to yellow (neutral performance) to green (profits), capped at -60% decrease and +60% increase per year. Therefore, the color coding of the performance values \( PM^A_{\text{annual}}(s, h) \) can be defined as a function

\[ \text{colormap}_{PM^A_{\text{annual}}}(s, h) : PM^A_{\text{annual}}(s, h) \to \text{Color}, \quad s, s-h \in T \]

that maps the performance values of all elements in the performance matrix into the corresponding color space.
3.2. Examples

Two examples where we applied our visualization technique on real-world data are shown in Figure 4. Figure 4(a) shows the performance of a ‘dot-com’ technology fund that we already used as introductory example in chapter 2.2. It is easy to derive from our box model that if sold in the year 2000 or before, this asset had been a good investment for all possible holding periods from 0 to 5 years, as nearly the complete corresponding area in the image is in dark green which reflects a strong growth of this asset.

It can be observed that large increases or decreases in the price of an asset in short periods of time are reflected by vertical and diagonal structures inside the image. The high increase in the price of the technology fund in Figure 4(a) in the beginning of the ‘dot-com’-boom in 2000 results in very high performance values for all holding periods from 0 to 5 years as the price of this fund was significantly lower in the previous 5 years, which is reflected by a nearly vertical dark green bar in the middle of the visualization.

However, it can also be seen that the image contains diagonal structures as well. The big loss in the price of this asset in January 2001 (the ‘time of the crash’), which therefore creates diagonal structures to all times of sale and holding times that refer to the date of the crash in January, 2001.

As mentioned, large impacts in the price produce vertical and diagonal structures, but as also can be seen in Figure 4(a), the diagonal green structure is evened out by the red vertical structure from the day of the crash and therefore ‘neutralized’. We can further identify short-term variability for holding times less than 3 months close to the x-axis by quickly changing red and green pixel areas that originate from the monthly ups and downs of the price.

As a temporary summary, we can conclude that our visual analysis technique allows a simultaneous comparison of all possible times of sale and all possible holding periods of an asset in one space efficient view, allowing an extensive analysis of an asset and offering better insight and understanding of the long-term characteristics of an asset compared to the statistical standard performance measures.

4. Enhancement by Relevance Functions

As statistical 1-year, 3-year and 5-year performance values only depict single points, an analyst is more often interested in having visual emphasis on the whole region surrounding a point of interest, as well as specifying multiple regions of interest with more complex relevance functions for interactively exploring the characteristics of an asset (or even groups of assets), yet still having a slight overview of the
complete set of data. In this chapter, we describe how we can extend our technique to be a relevance driven visualization that allows the investor to solve exactly these kind of problems.

Figure 5: The Weight Matrix WM shows the weights of all 6000 parameters, using a constant function for the relevance of the time of sale, and a Gauss function for the relevance of the holding period. The region of interest is set to a holding period of 3 years. The variance of the Gauss function controls the weight for holding periods longer and shorter than 3 years, which in the example still have a considerable weight.

When analyzing regions of interest in the performance of an asset, the analyst is typically interested in how an asset has performed for different holding times for a fixed time of sale (most often for the current time as the time of sale), or how an asset has performed over time for a fixed holding time [DG04]. In order to set a focus on the regions of interest, our application allows to model these user preferences with two relevance functions that can be easily modified using a GUI with multiple sliders. In order to set a focus on a region of interest, the investor can modify the relevance function $R_{Sale}$ to emphasize a region on the x-axis for an arbitrary time of sale, and $R_{Holding}$ to emphasize a region on the y-axis for a specific holding time (see Figure 5).

Several mathematical functions are available in the framework for specifying a relevance function, from constant over triangular to Gauss functions, where the investor can modify the parameter $(\mu)$ for the peak of the Gauss curve and the standard deviation $(\sigma)$ to modify the degree of uniformness of the relevance, according to his preferences.

Both functions $R_{Sale}$ and $R_{Holding}$ are then combined by

$$WM(s,h) = R_{Sale}(s) \cdot R_{Holding}(h)$$

for each possible time of sale $s$ and for each holding period $h$ to form a weighting matrix $WM(s,h)$, reflecting the investors’ preferences (see Figure 6).

For each of the two axis, two functions can be combined to create two regions of interest per axis, which (in case of the two Gauss functions as can be seen on the right in Figure 7) are simply added as

$$Sum = w_1 \cdot \frac{1}{2 \cdot \pi} \cdot e^{\frac{-(\mu-s)^2}{\sigma^2}} + w_2 \cdot \frac{1}{2 \cdot \pi} \cdot e^{\frac{-(\mu-h)^2}{\sigma^2}}$$

to build advanced weighting matrices that are capable of highlighting multiple regions of interest.

To visualize the weights of the relevance matrix seen in Figures 5 to 7, the normalized values $WM_a(s,h)$ are scaled onto a grayscale colormap according to

$$Colormap_{WM_a}: WM_a(s,h) \rightarrow Color$$

with a linear min/max normalization. The smallest value of the Weight Matrix $\min(WM_a)$ corresponds to white, whereas the highest value $\max(WM_a)$ of our matrix corresponds to black.

As the final step, the Weight Matrix $WM(s,h)$ is applied on the Performance Matrix $PM_a(s,h)$ from chapter 3 to get a realistic impression of the performance of an asset with focus on the investors region(s) of interest. The values of the Weight Matrix $WM(s,h)$ are used to modify saturation and brightness of the color values implied by the Performance Matrix $PM_a(s,h)$, intensifying the regions of interest and fading out the non-interesting areas.

As a simple example of how we can emphasize a special region of interest, we applied the relevance function in Figure 6 on the Performance Matrix in Figure 4(b), and get the result shown in Figure 8 where the investor sets a region of interest for a holding time of 3 years. We will demonstrate the advantage of this approach in the following chapter.
5. Analyzing Assets using the Weighted Performance Matrix

In the last decades, economic science has developed a wide variety of techniques with many different strategies to compare funds to one another for strategic asset allocation. Well known techniques for portfolio optimization and risk management are stochastic dominance, expected utility, value at risk, or the Arrow-Pratt measure of risk aversion. An in-depth discussion of the mathematical and stochastic algorithms from an economic point of view is out of the scope of this paper [D05]. However, we will show that these techniques can be significantly enhanced by our visual data analysis techniques.

As an example of how our visual techniques can be applied, we extend the well-known performance/risk analysis technique by applying our relevance functions from the previous chapter on it in order to get improved performance and risk measures that also take the investors special regions of interest and investment horizons into account. To obtain the improved performance scalar, we first normalize $W M_n$ in a way that the sum of all weights of $W M_n$ equals 1, by

$$W M_n(s, h) = \frac{W M(s, h)}{\sum_{s=s_{\min}}^{s_{\max}} \sum_{h=h_{\min}}^{h_{\max}} W M(k, l)}$$

In a second step, we compute the annualized values of our performance matrix $P M_a$ with the relevance preferences of the weighting matrix $W M_n$ according to

$$P M_a = E(P M_a) = \sum_{s=s_{\min}}^{s_{\max}} \sum_{h=h_{\min}}^{h_{\max}} W M_n(s, h) \cdot P M_a(s, h)$$

where $s$ represents the possible times of sale (the horizontal size of our performance and relevance matrix) and $h$ represents the possible holding periods (the vertical size of our matrix).
The risk of an asset, often also referred to as variance or volatility, can be computed in several different ways [D05]. The most often used method is the standard deviation according to

\[
\text{Variance}(X) = E(X^2) - (E(X))^2
\]

For calculating the risk, we again take the analysts preferences for special regions of interest into account, and combine the computation of the risk scalar by variance with our weighting matrix according to

\[
R_{PM^s} = \sum_{x_{\min}}^{x_{\max}} \sum_{h_{\min}}^{h_{\max}} W_{M^s}(s,h) \cdot (PM^s_h(s,h) - E(PM^s_h))^2
\]

We can now use these improved performance and risk scalars to compare the entire distribution with our own weighting for an advanced analysis.

5.2. Dominance Plots

An investor, who can choose from two assets with the identical risk, will typically choose the asset that promises a higher performance. Now that we can determine the performance and risk scalars \(P_{Bond}\) and \(R_{Bond}\) for all assets in our database for a given relevance function, it is possible to compare and to visualize the relationships that the assets have in order to have an overview of the distribution of performance and risk. Typically, an asset \(A\) and another asset \(B\) can differ in the way that one has a higher or lower performance than the other, or a higher or lower risk. The relationship (rel) between two assets \(A\) and \(B\) is one of the following:

- \(rel_{dom}: (\langle PA > PB \rangle \land (RA \leq RB) \lor (PA \geq PB) \land (RA < RB)) \rightarrow A\) dominates \(B\) (superiority)
- \(rel_{dom^s}: (\langle PA < PB \rangle \land (RA \geq RB)) \lor (PA \leq PB) \land (RA > RB)) \rightarrow A\) dominated by \(B\) (inferiority)
- \(rel_{higher}: (\langle PA > PB \rangle \land (RA > RB) \rightarrow A\) has more performance and more risk than \(B\) (no superiority or inferiority)
- \(rel_{lower}: (\langle PA < PB \rangle \land (RA < RB) \rightarrow A\) has less performance and less risk than \(B\) (no superiority or inferiority)
- \(rel_{equal}: (\langle PA = PB \rangle \land (RA = RB) \rightarrow A\) and \(B\) are equal.

where \(P\) is the performance scalar, and \(R\) is the corresponding risk scalar.

The relationships of the assets can be visualized in a Dominance Plot, giving an overview of where our reference asset is located compared to other assets in the database, and by how many other assets it is dominated and how many assets it dominates (see Figure 9).

A typical selection task by financial analysts is to find assets that are not dominated by any other asset, the so-called Pareto efficient assets. These assets have in common that no other asset exists in the data set that has less or similar risk but a higher performance, and are located on a concave function which is illustrated in Figure 10. The major advantage of our approach is that the assets on the efficiency curve can be located with respect to the investors regions of interest, and the icons offer an immediate insight into the characteristics of an asset with focus on the investors regions of interest.

6. Conclusions

In this paper, we have contributed a visualization technique for analysis of financial time series data that allows to have a complete overview on the characteristics of an asset for various holding times and times of sale. User-based relevance functions allow to emphasize single or multiple regions of interest and to focus on these areas for an advanced analysis. We have shown how our relevance functions can be used to compute improved performance and risk measures that also take the investors regions of interest into account, and how our visual Performance Matrix model can be integrated into traditional financial analysis techniques like Dominance Plots and efficiency curves to improve the decision making process by visualizing the detailed characteristics of assets.

Acknowledgements

This work was partially funded by the German Research Foundation (DFG) under grant GK-1042 ”Explorative Analysis and Visualization of Large Information Spaces”.

© The Eurographics Association 2007.
Figure 10: Assets that are not dominated by any other asset have in common that no second asset exists that offers a higher performance with a lower or identical risk. These assets form the Pareto efficiency frontier. It can be observed that the most efficient assets have low risk and low performance values, and look visually similar. By emphasizing the regions of interest in the icons, the investor has an instant overview of the development of an asset over time. For the given relevance matrix, we have a total of 12 assets that are not dominated by any other asset. Note that the Performance Matrices also provide a detailed view of the risk distribution for each asset (increasingly red regions from left to right).

References


[H03] HOCHHEISER, H.: Interactive Graphical Querying of Time Series and Linear Sequence Data Sets PhD. Dissertation, University of Maryland, Dept. of Computer Science, May 2003


[MS] MORNINGSTAR.: http://www.morningstar.com


[OV] ONVISTA MEDIA GMBH.: http://www.onvista.com


© The Eurographics Association 2007.