

Visual Analytics on the Financial Market: Pixel-based Analysis and Comparison of Long-Term Investments

Hartmut Ziegler Tilo Nietzsche Daniel A. Keim
University of Konstanz, Germany Credit Suisse, Switzerland University of Konstanz, Germany
ziegler@inf.uni-konstanz.de tilo.nietzsche@credit-suisse.com keim@inf.uni-konstanz.de

Abstract

In this paper, we describe solutions how pixel-based visualization techniques can support the decision making process for investors on the financial market. We especially focus on explorative interactive techniques where analysts try to analyze large amounts of financial data for long-term investments, and show how visualization can effectively support an investor to gain insight into large amounts of financial time series data. After presenting methods for improving the traditional performance/risk computation in order to take user-specific regions of interest into account, we present a novel visualization approach that demonstrates how changes in these regions of interest affect the ranking of assets in a long-term investment strategy.

Keywords—Visual Data Mining, Financial Data Analysis

1 Introduction

The financial market with its large amounts of time series data is an interesting and important domain for investors as well as for researchers. Although many techniques have been proposed how to gain insight into the data by using statistical or visualization methods, reliable knowledge discovery and decision support in this field is still a challenge and of high demand. Many techniques have been proposed to analyze or visualize the financial market as a whole, or single assets in particular, but still for many typical tasks of analysts the current solutions are not sufficient. In this paper, we focus on the visual decision support for long-term investments, and show a pixel-based paradigm for visual analysis of the long-term behavior of assets, and how we can use the regions of interest of an investor to modify the view on the data in order to find the assets that best suit specific long-term investment plans. Afterwards, we contribute a novel technique that shows how a change in the regions of interest affects the ranking of assets, and which assets ascend and which descent in the ranking of long-term investments.

2 Related Work

Large time series data in financial markets offer many possibilities for developing improved visualization techniques in order to get more insight into the market behavior. While the potential of using modern visualization techniques in the financial sector is high, many analysts or financial newspapers still rely on traditional line charts which are widely known and easy to understand for making their decisions. Unfortunately, information visualization is only slowly entering this domain, but a variety of approaches has been made in the recent years that have tried to visualize the financial market, for example with maps like treemaps [18], Neovisions RiskMaps [13], Smartmoneys 'Map of the Market' [16] or Self-Organizing Maps [1] [15] [4]. These techniques offer an adequate overview over the financial market, but do not allow any detailed analysis or comparison of single assets, neither analysis of effectively more than one interval of time and therefore cannot replace traditional line charts. Other applications that have been commercialized and that visualize financial time series data are GSphere by Gravity Investments [5], Market Topology [11], Portfolio Impact by High Tower Software [8], or StockVIS [14]. One project that is continuously improved is TimeSearcher [6], and another known project is ThemeRiver [7], which are not restricted to analyzing financial time series data but can be used with any time series data. One of the most recent approaches for visualizing financial time series data is based on pixel-based triangular matrices [9], which allow analysis of the development of an asset in overview and detail for one or many assets. Except the last, none of the other approaches allows the visualization of large amounts of time periods, and therefore do not provide sufficient information for an analyst for making reasonable investment decisions. In addition to that, other visualization techniques do not support to set a specified region of interest for detailed analysis of long-term investments while showing the growth rates for all possible time intervals between time of purchase and time of sale at the same time.

3 A Pixel-based Visualization Concept for the Analysis of Long-Term Investments

3.1 Shortcomings of Traditional Measures

Common traditional techniques for statistical financial data analysis often have the problem that their results are misleading, and can lead to misinterpretation with costly consequences. As a simple example, considering the 1-year, 3-year and 5-year performances of an asset, these statistical values only reflect the performance of three very specific time periods, relative to the current date. These parameters completely skip long periods of time which are situated in between, and skip 2 years of data where significant development is left out. We illustrate this problem in Figure 1 with an 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. 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, or the risk of the loss afterwards. If a second example is taken, and September 2001 happens to be the reference date for the 1-, 3- and 5-year performance, the three performance values would be completely different. Minimal changes in the holding period or a slightly different time of sale may cause the statistical parameters of volatile assets to change significantly by coincidentally hitting local peaks or lows.

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 measures, and may provide inadequate and unreliable information about the characteristics of an asset which makes them inappropriate for making important investment decisions.

3.2 The Performance Matrix Approach

One of the main interests for financial investments is the long-term performance of an asset. To overcome the mentioned problems which result from the lack of completeness by relying on only 3 out of many thousand possible time intervals, we proposed a pixel-based visualization paradigm named *Performance Matrix* [19]. This two-dimensional rectangular approach, as shown in Figure 2, allows simultaneous comparison of 4500 combinations of time of sale and holding period in one space efficient pixel-based view, forming a Matrix PM with 90 degree angles and giving a much deeper, more reliable and easier to interpret insight into the long-term characteristics of an asset. In

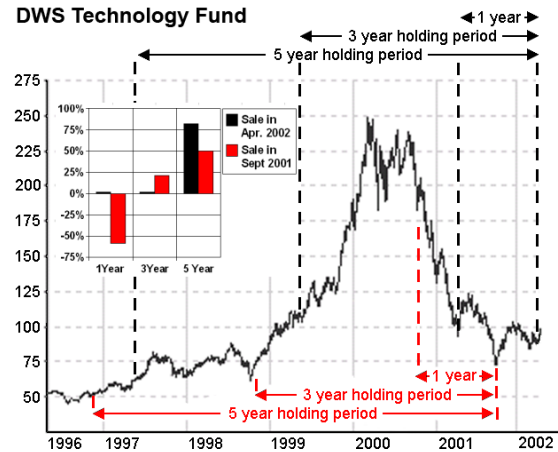


Figure 1: Inadequacies of standard performance ratios. Price chart illustrating a technology fund during the 'dot-com' crisis. The financial ratios for 5-year, 3-year and 1-year performance do not represent the characteristics of the asset because the peak in the year 2000 is completely situated between the time intervals. Shifting the reference point slightly backwards the ratios return completely different values although most of the chart is identical.

order to better illustrate this principle, the statistical measures $PM(2005, 1y)$, $PM(2005, 3y)$ and $PM(2005, 5y)$ for 1, 3 and 5 year performance are denoted in Figure 2. The performance for each combination of holding period and time of sale can be described as a metric $PM^A(s, h)$:

$$PM^A(s, h) = \frac{V^A(s)}{V^A(s-h)} \quad s, s-h \in T$$

where s is the time of sale, $s-h$ is the time of purchase of an asset A taken from the set of all possible points of time T , and h is the holding period. Each pixel is colored corresponding to the growth rate of the asset for this time interval. We use the color map in Figure 3 for coloring the values. Therefore, the problem of having incomplete statistical values that do not describe the characteristics of an asset sufficiently can easily be solved with a space-efficient visualization technique.

Two examples can be seen in Figure 4. While the bond with its continuous moderate growth rate is a stable and risk-free long-term investment for all possible holding periods, the technology fund has been a good long-term investment until the year 2000, but even to the year 2002 if the asset has been bought before 1998. The short-term volatility of the asset can be recognized by the quickly changing red and green pixels directly along the x-axis. This way, each asset generates its own fingerprint that we can use to visually analyze the long-term behavior of each asset and to gain insight into its characteristics.

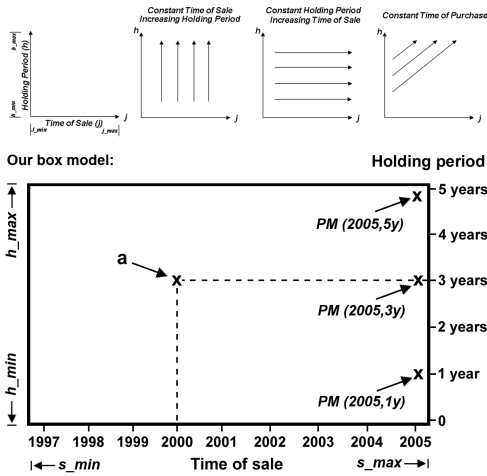


Figure 2: This figure shows the rectangular *Performance Matrix* model for analyzing characteristics of long-term investments. The x-axis represents all times of sale and the y-axis represents all holding periods. If an asset has been a good 3- or 5-year long-term investment can be analyzed by the corresponding horizontal line. The point "a" for example denotes the 3-year-performance of an asset with a time of sale in January 2000.

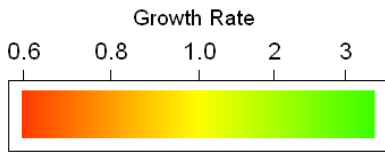


Figure 3: Our colormap visualizes the performance of an asset with a linear spectrum from red (loss) to yellow (neutral performance) to green (profit).

4 Modeling of Individual Preferences

As statistical 1-year, 3-year and 5-year performance values only depict single points, an analyst is more often interested in also having visual emphasis on the whole region surrounding a point of interest. 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. 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 (see Figure 5). 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

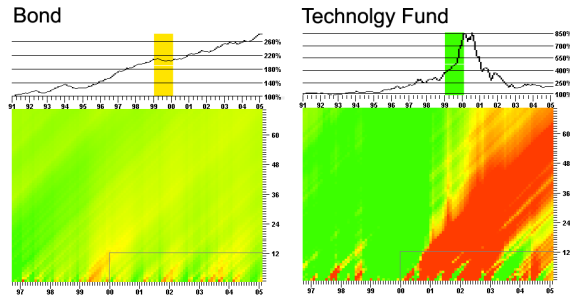


Figure 4: Two example assets show how the lack of incomplete statistical values can be compensated by a pixel-based visualization technique. The bond with its continuous growth is a good investment in the long term for almost all holding times as most of the matrix is green, while the technology fund clearly reveals the problems of this asset in the last five years.

y-axis for a specific holding time.

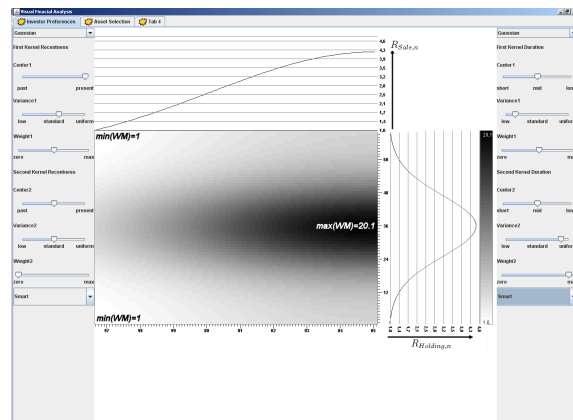


Figure 5: The Weight Matrix (WM) shows the weights of all combinations of time of sale and holding period, using Gauss functions to model the relevance on each axis for the time of sale and the holding period. To emphasize recent price movements the peak of the Gauss function on the x-axis focuses on the current date, the peak on the y-axis on a holding period of 3 years.

As the final step, the Weight Matrix $WM(s, h)$ is applied on the Performance Matrix $PM_a^A(s, h)$ 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 $PM_a^A(s, h)$, intensifying the regions of interest and fading out in the non-interesting areas. As example of how we can emphasize special regions of interest for long term investments, we apply the weight function for the four scenarios in Figure 6 on the Performance Matrix in Figure 4(b), and get the result shown in Figure 7.

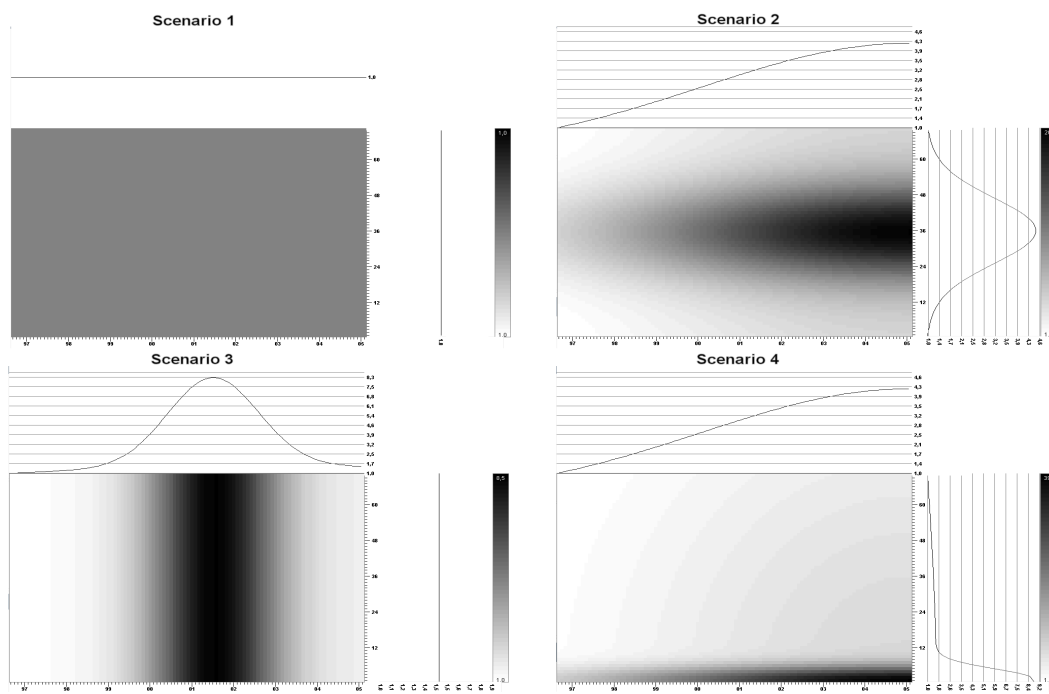


Figure 6: In the first scenario the analyst does not focus on any certain investment periods or times of sale. The second scenario shows a typical private investor being mainly interested in a holding period of 3 years and being more interested in the recent development of an asset. The third scenario focuses on the dot.com phenomenon while the fourth scenario focuses on the short term behavior.

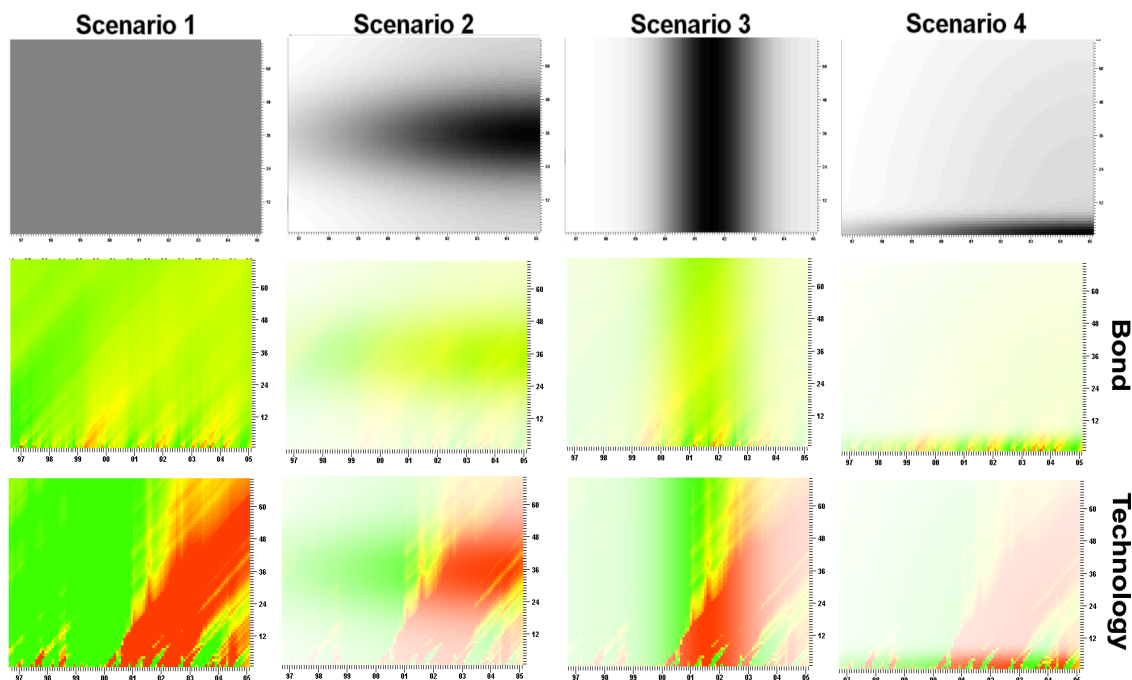


Figure 7: Visualization of four analysis scenarios using the weighting functions on our Performance Matrices in Figure 4(a) and (b). The spotlight effect on the region of interest creates a personalized representation of the bond fund (*DWS Eurenta Fond*) and the technology fund (*DWS Technologie Fond*). This instantly reveals which fund is a good investment for which particular investment time.

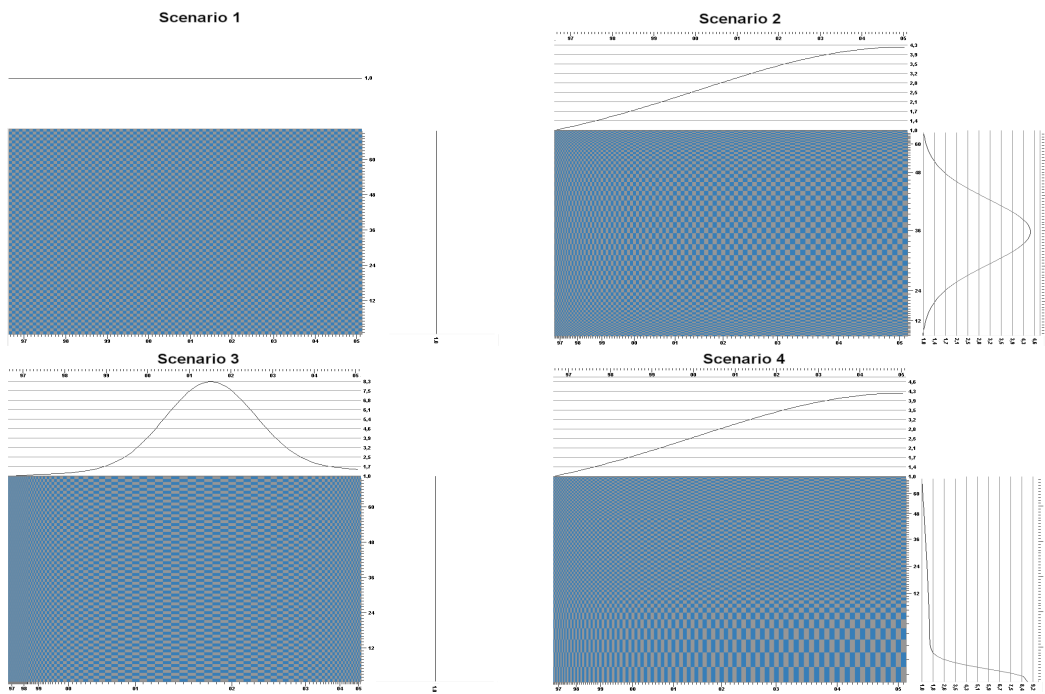


Figure 8: Four scenarios for visualization of long-term investments using distortion. The areas of interest of the investor are enlarged, whereas regions with minor interest shrink, therefore the map is proportionally partitioned according to the interest of an investor.

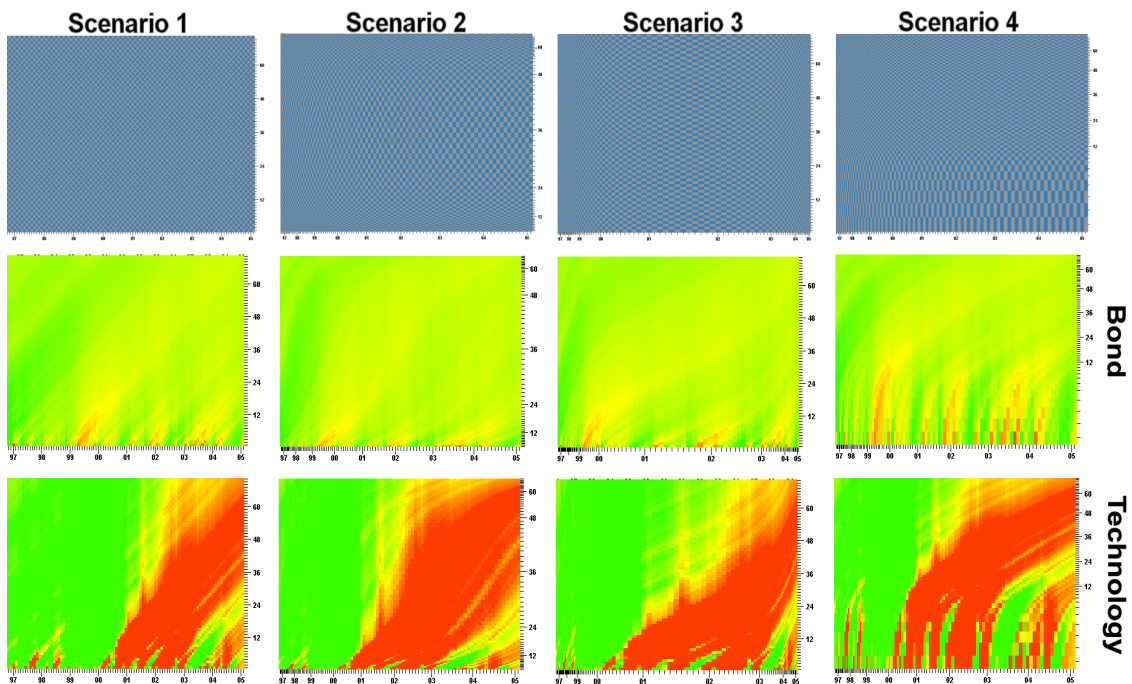


Figure 9: If the four scenarios from figure 8 are applied on our two Performance Matrices in Figure 4(a) and (b), the images support the decision making process by offering an optimized view on ones individual regions of interest. It can be identified that the bond with its solid long-term characteristics still has its risk in the short-time volatility, or the technology fund offers some risky but highly interesting short-time investment opportunities.

Instead of reflecting the importance value by a modification of the color, an alternative approach distorts the space that is used for the representation. Important regions of interest get larger areas, whereas unimportant investments get smaller areas. Therefore, the area of the *Performance Matrix* is partitioned proportional to regions of interest (see figures 8 and 9). This technique allows to use more display space for the important information.

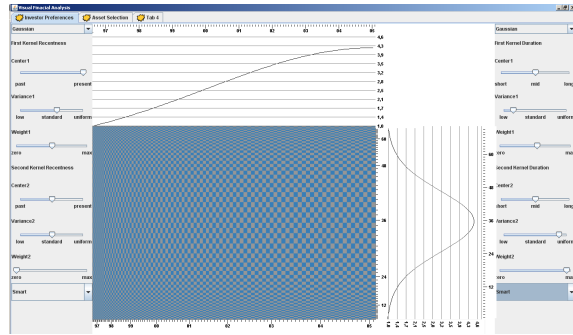


Figure 10: In this image, the interactive tool for setting the distortion according to the investors' region of interest focuses on the recent development with the main focus on 3 year investments.

5 Visual Performance/Risks Analysis

5.1 Enhanced Performance/Risks Scalars

A well known traditional statistical method for the evaluation of assets is the calculation of the risk (volatility) and performance. These two scalar values can be determined for each asset individually, and are usually computed without taking the investors regions of interest into account. In order to improve these scalars, instead of weighting all values identically, we developed enhanced risk and performance scalars that also take the regions of interest into account. Once these values have been computed and take the weight matrix into account, it is possible to directly compare assets one to each other with respect to these two scalar variables. Typically, an investor who has the choice between two assets, will take the asset which offers the higher performance with less or identical risk. So when comparing different assets and their scalars one to each other, five possible combinations can occur. Regarding the risk scalar (R) and the performance scalar (P) the relationship (rel) between two assets A and B can be

- $rel_{dom}: ((P_A > P_B) \wedge (R_A \leq R_B)) \vee ((P_A \geq P_B) \wedge (R_A < R_B)) \rightarrow A \text{ dominates } B$ (superiority)
- $rel_{isDom}: ((P_A < P_B) \wedge (R_A \geq R_B)) \vee ((P_A \leq P_B) \wedge (R_A > R_B)) \rightarrow A \text{ dominated by } B$

(inferiority)

- $rel_{higher}: (P_A > P_B) \wedge (R_A > R_B) \rightarrow A \text{ has more performance and more risk than } B$ (no superiority or inferiority)
- $rel_{lower}: (P_A < P_B) \wedge (R_A < R_B) \rightarrow A \text{ has less performance and less risk than } B$ (no superiority or inferiority)
- $rel_{equal}: (P_A = P_B) \wedge (R_A = R_B) \rightarrow \text{Equality}$

A typical task by financial analysts is to find assets that are not dominated by any other asset. These assets have in common that no other asset with a higher performance exists which has less or equal risk. The left diagram of Figure 12 shows the so called *efficient assets* using the investors' preferences from scenario 2. The right diagram displays the relationship of these 16 assets to all other assets. An advantage of this analysis is that assets on the efficiency curve are positioned with respect to the investors' regions of interest. The Performance Matrix icons allow a full visual analysis of the corresponding assets in one view.

5.2 Comparative Analysis of Multiple Preferences

In this section, we present a new decision support visualization which displays the relationships of all assets while still allowing in-depth analysis of a single asset. Regarding the five possibilities in the previous section, we can not only determine how attractive an asset is compared to another asset, but can also further analyze the changes that occur when the investor focuses on a different region of interest and changes the weight matrix. For identifying interesting investment objects, the analyst can filter for efficient assets and compare the changes between different preferences. The idea of this analysis is to identify assets that benefit and assets that lose by changing the scenario (see Figure 6). If several investors use the first scenario and another investor is interested in the second analysis scenario the assets that benefit from this change can be interesting investment objects whereas the losers will not be of interest. Figure 11 describes how changes of the relationships are treated. Winners and losers are determined by summing up all *change values* per asset. The maximum value corresponds to the top gainer and the minimum value to the top loser asset. Figure 13 shows two scenarios and the resulting relationship of the assets. Integrating the migration matrix from Figure 11 the differences can be pointed out and the losers and gainers can be identified in order to support the analyst in finding potentially interesting investment objects.

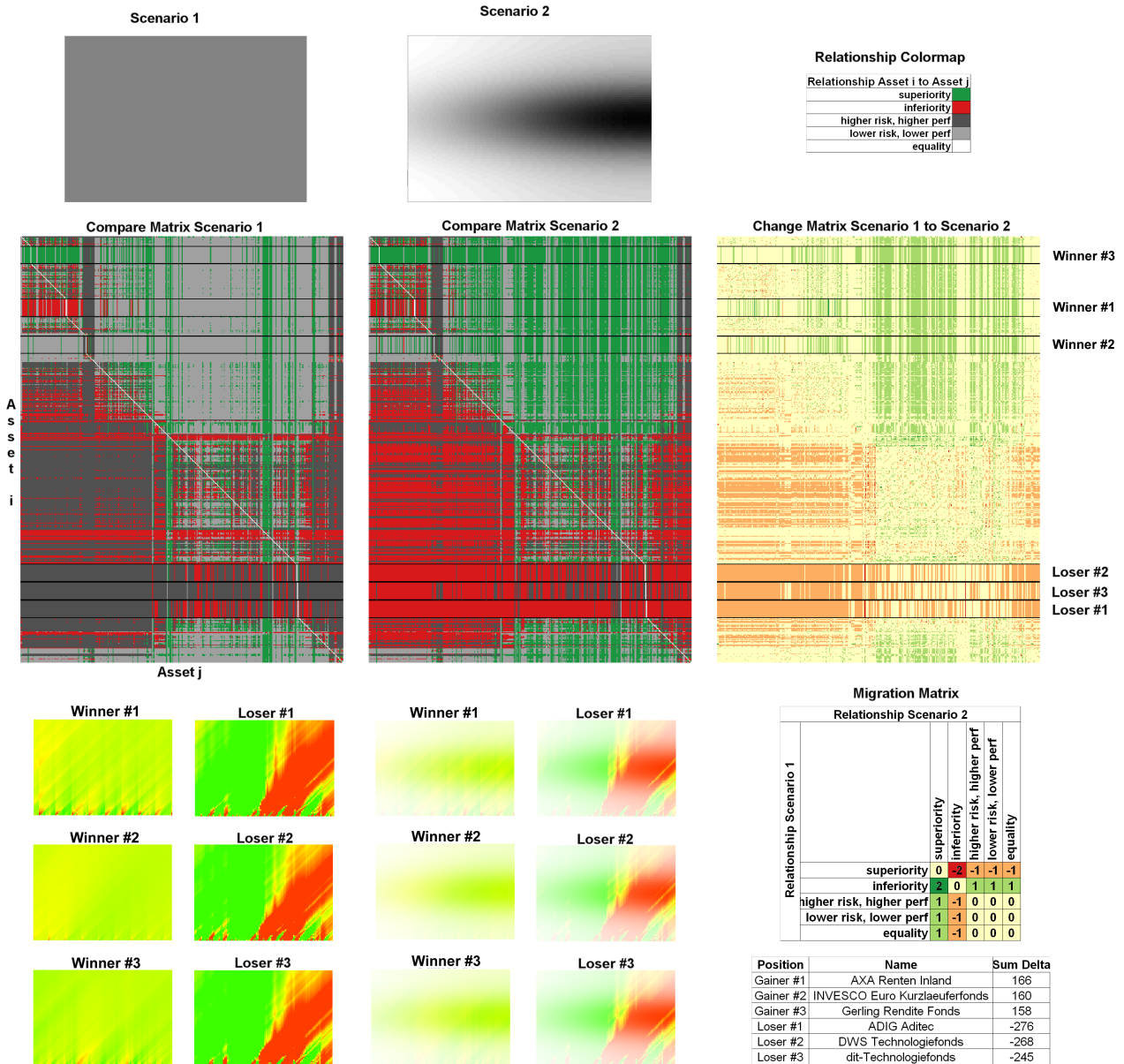


Figure 13: Analyzing winners and losers resulting from a change from scenario one to scenario two. The compare matrices display the relationships of all assets for the scenarios one and two. With the integration of the migration matrix the change matrix can be constructed. The change matrix and the compare matrices highlight the top three gainer and loser assets. The lower part of the figure displays the Performance Matrices of these assets for both scenarios. The winners profit because the weights of short term behavior and early times of sale are reduced. In these areas the winners show a sideways movements. The losers suffer from reducing these weights as they made high profits in the corresponding areas and high losses in the areas that are of interest regarding the second scenario.

6 Future Work

This research will be continued in several directions. First, a usability study has to be conducted which quantitatively determines the benefits of the proposed techniques

and which determines further improvement possibilities regarding colormaps and interaction techniques. Due to the fact that this is a special visualization technique developed to support financial experts, the candidates for the usability study must also be taken from this group. Furthermore, representative assets can be determined based on visual

Relationship Scenario 1		Relationship Scenario 2				
		superiority	inferiority	higher risk, higher perf	lower risk, lower perf	equality
superiority	0	-2	-1	-1	-1	
inferiority	2	0	1	1	1	
higher risk, higher perf	1	-1	0	0	0	
lower risk, lower perf	1	-1	0	0	0	
equality	1	-1	0	0	0	

Figure 11: The migration matrix shows colors and change values that are assigned resulting from a change of the scenario. For example, if asset i is superior to asset j in the first scenario and i is inferior to j in the second scenario then a change value of -2 and the red color is assigned. Summing up all change values per asset enables to determine the gainers and losers of the change.

similarity measures. The work could further be extended by pre-computing performance and risk scalars to speed up the application.

Conclusion

Our methods show how modern visualization techniques can effectively extend the knowledge discovery and decision making process in financial markets for analysts and investors, with valuable insights that usually cannot be easily derived with any other traditional or existing visualization method. The Performance Matrix, a pixel-based visualization technique, features a structure that focuses on the analysis of assets with respect to multiple holding periods and times of sale. Integrating the individual preferences of an investor, personalized fingerprint visualizations can be created that visually reflect the regions of interest by putting emphasis on these regions using a spotlight effect or enlarging these regions using distortion, and improved performance and risk scalars can be computed according to these weights. This concept leads to fast decision support while still offering a detailed insight into the selected assets for all time intervals. Filtering for efficient assets by using efficiency curves and analyzing the changes between two scenarios with different individual user preferences allows identification of interesting investment objects and improves insight into the large amounts of data in the financial domain.

Acknowledgements

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

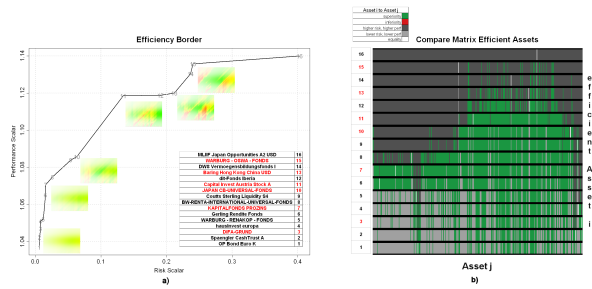


Figure 12: a) Efficient assets regarding the second scenario with a selection of the 16 efficient assets. By emphasizing the regions of interest in the icons, the investor has an instant overview on the past development of these assets. The displayed Performance Matrices provide a detailed view of the risk distribution for each asset (increasingly red regions from left to right). b) The right diagram shows the relationship of the 16 efficient assets to a set of 392 other assets in the DB. The efficient assets are ordered by increasing risk values from the bottom to the top. Efficient assets with low risk often have the relationship rel_{lower} indicated by the brighter grey whereas efficient assets with higher risk often have the rel_{higher} relationship indicated by darker grey.

References

- [1] Barbro Back, Mikko Irjala, Kaisa Sere, Hannu Vanharanta, *Competitive Financial Benchmarking Using Self-Organizing Maps*. Department of Computer Science, Series A No 169, Abo, 1995
- [2] Kevin Dowd, *Measuring Market Risk.*, Wiley and Sons, 2005, 2nd Edition
- [3] Tim Dwyer, David R. Gallagher. *Visualising changes in fund manager holdings in two and a half-dimensions.*, in *Information Visualization 3, 4* (2004), 227–244.
- [4] Guido J. Deboeck, Tuevo K. Kohonen. *Visual Explorations in Finance with Self Organizing Maps*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1998
- [5] Gravity Investments, Inc.: GSphere, http://www.gravityinvestments.com/product_tour.htm
- [6] Hochheiser, Harry; *Interactive Graphical Querying of Time Series and Linear Sequence Data Sets*, PhD. Dissertation, University of Maryland, Dept. of Computer Science, May 2003
- [7] Havre, Susan, Hetzler, Elizabeth G., Nowell Lucy T.: *ThemeRiver: Visualizing Theme Changes over Time*, Proceedings of the Infovis 2000, p. 115-124
- [8] High Tower Software, Inc.: Portfolio Impact, <http://www.portfolioimpact.com/>
- [9] Daniel A. Keim, Tilo Nietzsche, Norman Schelwies, Jörn Schneidewind, Tobias Schreck and Hartmut Ziegler. *A Spectral Visualization System for Analyzing Financial Time Series Data*. In *Proceeding EuroVis*

2006: Eurographics/IEEE-VGTC Symposium on Visualization (EUROVIS'06), Lisbon, Portugal, May 2006

- [10] Jessica Lin, Eamonn Keogh, Stefano Lonardi, Jeffrey P. Lankford, Donna M. Nystrom. *Visually mining and monitoring massive time series*. In *KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (New York, NY, USA, 2004), ACM Press, pp. 460–469.
- [11] Market Topology: url: <http://www.market-topology.com>
- [12] John J. Murphy (Ed.). *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*. Prentice Hall Press, 1999.
- [13] NeoVision Hypersystems Inc.: *HeatMaps and RiskMaps*, url: <http://www.neovision.com>
- [14] Edward Parrish. *StockVis: An Internet-Based System for Visualizing Stock Market Data*. Master Thesis, UC Santa Cruz, Department of Computer Science, 2000
- [15] Kresimir Semunic, *Visualizing of Stock Market Charts*. In Proceedings from the 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision 2003 (2003), Plzen-Bory (CZ), 2003
- [16] Smart Money: 'Map of the Market', <http://www.smartmoney.com/marketmap/>, 2005
- [17] Heidrun Schumann, Wolfgang Mueller. *Visualization methods for time dependent data - an overview*, In Proc. Winter Simulation Conference (2003), ACM, pp. 737–745.
- [18] Martin Wattenberg. *Visualizing the stock market*. In CHI '99: CHI'99 extended abstracts on Human factors in computing systems, New York, NY, USA, 1999. ACM Press, pp. 188-189.
- [19] Hartmut Ziegler, Tilo Nietzsche, Daniel A. Keim. *Relevance Driven Visualization of Financial Performance Measures*, In Proceedings of the EuroVis2007: Eurographics/IEEE-VGTC Symposium on Visualization (EUROVIS'07), Norrköping, Sweden, May 2007
- [20] Hartmut Ziegler, Tilo Nietzsche, Daniel A. Keim. *Visual exploration and discovery of atypical behavior in financial time series data using two-dimensional colormaps*, In Proceedings 11. International Conference on Information Visualization (IV '07), 2007, pp. 308-315