Dispersion of Returns
An Analysis in Emerging Markets

Ashwin Goyal, Thejes Venkatesh, Yu Wang, Ben Xie
Pratt School of Engineering, Duke University

BACKGROUND

INSTITUTIONAL INVESTORS STAND TO PROFIT IF THEY CAN PREDICT PERIODS PERIODS OF HIGH CROSS-SECTIONAL VOLATILITY

Cross-Sectional Volatility, or dispersion, is defined as the standard deviation of returns over a given period. Zero dispersion occurs when all active portfolio returns equals the index return. Thus skilled active investors perform best in environments with high dispersion.

The ability to predict periods of high dispersion will allow institutional investors to optimize the allocation of money for active management.

The methodology to predict periods of high diserpsion was explored using data provided by Russell. The dataset consisted of hundreds of thousands of emerging market asset values.

Once the dataset was processed a variety of methods were used to predict dispersion
1. Autocorrelations and correlations of key descriptive statistics were calculated to be used as potential signals for high dispersion
2. Pattern recognition of morphological trends in the histograms of monthly returns were analyzed to determine if future dispersion was related to factors in current dispersion

DATA PROCESSING

THE RUSSELL DATASET NEEDED TO BOTH CONVERTED INTO A USEABLE FORMAT AND PROCESSED TO REMOVE ERRONEOUS DATA

We were provided with a huge dataset which had emerging markets stock data for 13 years starting from July 1996 to August 2009. The data represented 50 countries, with 5801 unique stocks. The number of stocks in each of the 157 months varied between 700 and 2200- in all 277,673 data points.

The shares were either value shares or growth shares-some were both. The challenges we faced with the data were mainly of two hues- 1.The data was not organized 2. It did not have the attributes we needed for our analysis-i.e. Stock returns. We first converted the messy data into tabular form and calculated the share price by dividing the market capitalization by the number of shares outstanding.

Sample of Original Data

```
AsOfDate InstrumentCode Description IndexFamily GrowthValueFlag
OneDayTotalPerfUSD OneDayNetPerfUSD TotalRetIndexUSD NetTotRetIndexUSD
OneDayTotalPerfEUR OneDayNetPerfEUR TotalRetIndexEUR NetTotRetIndexEUR
01/01/1997 R89950 Emerging Mkts AC C 0.000000 0.000000 1376.499550 1386.715608
NULL NULL NULL NULL 01/01/1997 R95114 Emerging Mkts Growth AC G 0.000000
0.000000 1887.863233 190.006705 NULL NULL NULL NULL 01/01/1997 R95144 Emerging
Mkts Value AC V 0.000000 0.000000 1065.299118 107.423137 NULL NULL NULL NULL
01/01/1997 R90125 Emerging Mkts LC LC C 0.000000 0.000000 1264.980841
1273.856552 NULL NULL NULL NULL 01/01/1997 R95173 Emerging Mkts LC Growth LC G
0.000000 0.000000 1827.468253 183.924473 NULL NULL NULL NULL 01/01/1997 R95201
Emerging Mkts LC Value LC V 0.000000 0.000000 929.180035 93.640703 NULL NULL
```

Russell tracks the free float market capitalization-which is the shares that are actually traded. And because it does not consider locked in shares such as those held by the government and promoters, it is seen as a more accurate indicator of market movements. Also, Russell applies market cap screens and knocks out the bottom 2% of the stocks every May and replaces it with new stocks. Hence there is fluctuation in the number of stocks in the months between May and July.

We faced further complexity in getting accurate stock returns because of stock splits and reverse splits. When splits happen, the market capitalization remains approximately the same but the number of stocks change-impacting the stock price. We fixed this issue using an algorithm to identify stock splits and calculate the returns based on market capitalization.

GENERAL RESEARCH GOALS

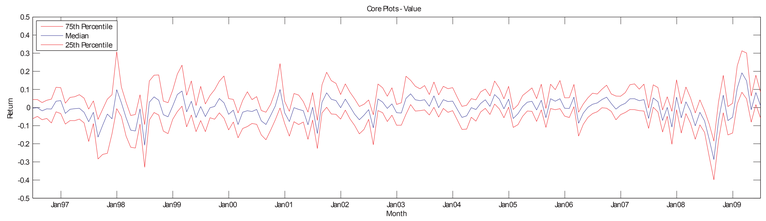
1. Develop a scalable protocol to automatically process large datasets and remove erroneous values

2. Extract and analyze key metrics from the data set

3. Determine a reliable method to signal for dispersion

Specific Aim #2

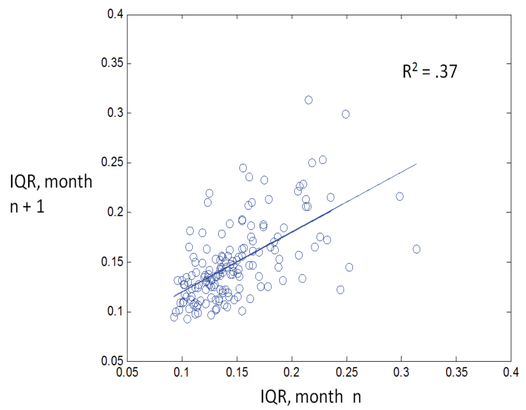
CALCULATE KEY DATASET STATISTICS



From the processed data, the 25th percentile, median, and 75th percentile asset return was calculated for each month through a sorting algorithm. These values were plotted through time for both the Growth and Value universes, and show general directional moves in the market. Cross sectional volatility can be represented by the distance between the 75th and 25th percentile returns (red lines).

Specific Aim #3

CALCULATE AUTOCORRELATIONS FOR GROWTH AND VALUE



In order to predict dispersion in future time periods, autocorrelation analysis was run on the dispersion values by running a least-squares linear regression for the IQR of month n against the IQR of month n+1. With significant R2 values of 0.28 and 0.37 for Value and Growth, respectively, it is clear that there is information contained about the dispersion of the next month within the market's current month dispersion. This is a significant signaling discovery, and provides an elementary basis for predicting future cross-sectional volatility of the market.

Specific Aim #4

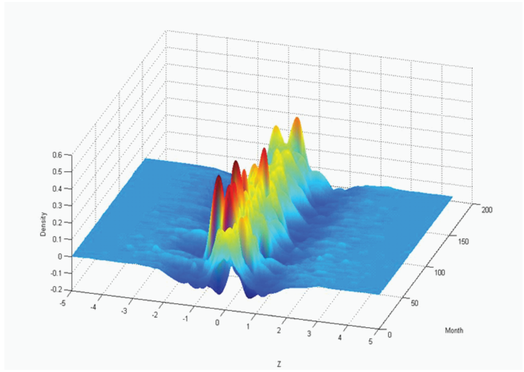
ANALYSIS OF PATTERNS BASED ON HISTOGRAMS OF MONTHLY RETURNS

Based on the stock return of each month, we normalized the data by using (stock return – mean of that month)/standard deviation of that month. The normalization provides us benefits to compare data from month to month.

Histograms provide general information on how the stock returns spread. Based on that, we use KDE1 (Kernel Density Estimator) function to estimate the probability density of the return distribution for each month. This method helps us smooth the historical data, prior to comparison with the normal curve.

To dig into the signals, we focus on the difference between smoothed histogram curve and normal distribution curve. Using histogram curve minus the normal distribution curve, we create the shape below.

Comparing curves of different months, we can clearly find persistence among the shapes from month to month. To visualize the curve movement of the entire period, we create the 3D pictures as below.



Month from 1 to 157 equals the month from Aug, 1996 to Aug, 2009 respectively. Z here equals Z score. The color shows the value intensity. This picture shows how the peak value moves along time and helps to locate their time position. Below is the same picture from the Z score perspective.

ACKNOWLEDGEMENTS

We sincerely thank Daniel Egger, Jai Jacobs, Aino Levonmaa, and Emma Rasiel for their invaluable project insight. Financial support was provided from Lazard and Russell

REFERENCES

1. Ankrim, E. and Ding, Z. "Cross-sectional Volatility and Active Manager Return Dispersion." Jun 2001, Russell Research Commentary.
2. Botev, Z.I. "A Novel Nonparametric Density Estimator". Nov 2006, University of Queensland.
3. Hwang, S. and Satchell, S. "Properties of Cross-sectional Volatility." Sept 2004, Case Business School.