Financial Applications of the Mahalanobis Distance
26th Australasian Finance and Banking Conference 2013

S. Stöckl¹, M. Hanke

Chair of Finance
Institute for Financial Services
University of Liechtenstein

Thursday, December 19th, 2013

¹Corresponding author; email: sebastian.stoeckl@uni.li; tel: +423 265 1153.
Many problems in finance involve multivariate random variables $\Rightarrow$ Similarity of realizations?
Many problems in finance involve multivariate random variables ⇒ Similarity of realizations?

⇒ Suitable (based on statistical theory) measure for detection: the Mahalanobis distance (MD)

Examples: outlier detection, portfolio surveillance, asset classification
Introduction

- Application based on MD: Financial Turbulence [Kritzmann & Li, 2010]
- Multivariate unusualness in financial market data

\[ FT_t = (r_t - \mu)' \Sigma^{-1} (r_t - \mu). \] (1)

- \( \Rightarrow \) Based on squared MD
Introduction

- Application based on MD: Financial Turbulence [Kritzmann & Li, 2010]
- Multivariate unusualness in financial market data

\[ FT_t = (r_t - \mu) \Sigma^{-1} (r_t - \mu). \]  \hspace{1cm} (1)

- \( \Rightarrow \) Based on squared MD
- \( r_t, \mu \) and \( \Sigma \) may each be determined or estimated in various different ways.
Introduction

- Application based on MD: Financial Turbulence [Kritzmann & Li, 2010]
- Multivariate unusualness in financial market data

\[ FT_t = (r_t - \mu)' \Sigma^{-1} (r_t - \mu). \]  

⇒ Based on squared MD
- \( r_t, \mu \) and \( \Sigma \) may each be determined or estimated in various different ways.
- ⇒ Depending on inputs, the resulting Mahalanobis distance will have a different economic interpretation.

Aim of the paper:

Explore promising combinations (Input).
Discuss previous and potential uses and usefulness for financial market participants (Output).
Contents

1 Introduction

2 Properties of the Mahalanobis distance

3 Types of Financial Applications

4 Conclusion
Initial Motivation for MD

- Analyze and classify human skulls into groups, based on different properties [Mahalanobis, 1927]
- Properties for group classification in a financial context:
  - Returns on assets in a portfolio
  - Portfolio properties for an investment company
Use of MD suggested by Kritzmann & Li (2010)

- MD as indicator for unusualness in financial markets (Financial Turbulence)
- Return of one asset in relation to it’s historical mean and standard deviation:
  \[(r_t - \mu)^2 / \sigma^2\]
  \[\Rightarrow (\text{Squared}) \text{ Mahalanobis Distance for one asset}\]
Use of MD suggested by Kritzmann & Li (2010)

- MD as indicator for unusualness in financial markets (Financial Turbulence)
- Return of one asset in relation to it’s historical mean and standard deviation: \((r_t - \mu)^2/\sigma^2\)
  \[\Rightarrow (\text{Squared}) \, \text{Mahalanobis Distance for one asset}\]
- Multivariate extension in a portfolio context

\[
\text{Eu}^2_t = \sum_{i=1}^{n} \frac{w_i^2}{w^2} \frac{(r_{t,i} - \mu_i)^2}{\sigma_i^2}
\]

\[\Rightarrow \text{Weighted, squared and normalized Euclidean Distance}\]

\[
\text{Ma}^2_t = \frac{w^2}{w^2} \frac{(r_t - \mu)^2}{\Sigma^{-1}_w (r_t - \mu)} \quad (3)
\]

\[\Rightarrow \text{Weighted, squared and normalized Mahalanobis Distance (Portfolio Turbulence)}\]

\[\Rightarrow \text{Additionally treats joint deviations of portfolio returns } r_i \text{ and } r_j \text{ given by their correlation } \rho_{ij}\]
Use of MD suggested by Kritzmann & Li (2010)

- MD as indicator for unusualness in financial markets (Financial Turbulence)
- Return of one asset in relation to it’s historical mean and standard deviation: 
  \[ (r_t - \mu)^2 / \sigma^2 \]
  ⇒ (Squared) Mahalanobis Distance for one asset
- Multivariate extension in a portfolio context

\[ E_{u_t}^2 = \sum_{i=1}^{n} \frac{w_i^2 (r_{t,i} - \mu_i)^2}{\sigma_i^2} \]  
  \[ \Rightarrow \] Weighted, squared and normalized Euclidean Distance

- Including information on the direction of moves

\[ M_{a_t}^2 = \frac{1}{w^2} (r_t - \mu)' w_D \Sigma^{-1} w_D (r_t - \mu) \]  
  \[ \Rightarrow \] Weighted, squared and normalized Mahalanobis Distance (Portfolio Turbulence)
  ⇒ Additionally treats joint deviations of portfolio returns \( r_i \) and \( r_j \) given by their correlation \( \rho_{ij} \)
(Statistical) Properties of the MD

- (Squared) MD has a $\chi^2(n)$-distribution (given $r_t \sim N_n(\mu, \Sigma)$)

- Normalization leads to an Expectation of 1

- MD is invariant under affine transformation $Y = a + BX$ [Meucci, 2009]

- MD captures all statistical information for elliptical distributions (fully described by location parameter $\mu$ and scatter matrix $\Sigma$)
(Statistical) Properties of the MD

- (Squared) MD has a $\chi^2(n)$-distribution (given $r_t \sim N_n(\mu, \Sigma)$)
- (Weighted & squared) MD has a generalized $\chi^2$-distribution with parameters $\Sigma$ and $w_D\Sigma^{-1}w_D$ and Expectation $w^2 = tr(w_D^2) = \sum_{i=1}^{n} w_i^2$
- Normalization leads to an Expectation of 1
(Statistical) Properties of the MD

- (Squared) MD has a $\chi^2(n)$-distribution (given $r_t \sim N_n(\mu, \Sigma)$)
- (Weighted & squared) MD has a generalized $\chi^2$-distribution with parameters $\Sigma$ and $w_D \Sigma^{-1} w_D$ and Expectation $w^2 = tr(w_D^2) = \sum_{i=1}^n w_i^2$
- Normalization leads to an Expectation of 1
- MD is invariant under affine transformation $Y = a + B \dot{X}$ [Meucci, 2009]
- MD captures all statistical information for elliptical distributions (fully described by location parameter $\mu$ and scatter matrix $\Sigma$)
Contents

1 Introduction

2 Properties of the Mahalanobis distance

3 Types of Financial Applications
   - Sample-based Differences
   - Deviations from Model Prices
   - Forecast Evaluation

4 Conclusion
Sample-based input parameters
To detect major changes in markets

Distinguish two types of sample based differences:
- Compare a realization to a historical sample or
- Compare two samples
Sample-based input parameters

Compare a realization to a historical sample

- Financial and Portfolio turbulence: relates today’s returns to historical means ($\mu = \bar{x}$) and sample covariances ($\Sigma = S$).

- $\Rightarrow$ Standardized indicator of unusual behavior across markets and portfolios.

- Useful for market/portfolio surveillance [Bodnar, 2009].
Sample-based input parameters

Compare two samples

- Compare two samples of different time periods by relating their means and covariance matrices against each other [McLachlan, 1999]
  \[ \Rightarrow \text{Hotelling } T^2 \text{-test for multivariate dependent samples, based on (squared) MD [Rao, 2009].} \]
Sample-based input parameters

Compare two samples

- Compare two samples of different time periods by relating their means and covariance matrices against each other [McLachlan, 1999]
  \[ \Rightarrow \text{Hotelling } T^2\text{-test for multivariate dependent samples, based on (squared) MD } [\text{Rao, 2009}]. \]

- Use both methods to determine periods of (non-) turbulence.

- Use turbulent market parameters for stress testing portfolios [Chow, 1999].

- Use non-turbulent market parameters (removing outliers) for robust portfolio estimation [Campbell, 1998].
Sample-based input parameters
Example 1: Change in market conditions during the financial crisis

- Portfolio equities (3/8), fixed income (2/8), real estate (2/8) and alternative investments (1/8)
- Historical sample calibration period: 2004-2006

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>std.dev</th>
<th>skewness</th>
<th>kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equities</td>
<td>0.0006</td>
<td>0.0058</td>
<td>-0.23</td>
<td>1.02</td>
</tr>
<tr>
<td>Fixed income</td>
<td>0.0001</td>
<td>0.0035</td>
<td>0.00</td>
<td>1.01</td>
</tr>
<tr>
<td>Real estate</td>
<td>0.0010</td>
<td>0.0067</td>
<td>-0.49</td>
<td>1.40</td>
</tr>
<tr>
<td>Alt. investments</td>
<td>0.0003</td>
<td>0.0146</td>
<td>0.13</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Table 1: Descriptive statistics - Time frame: 01/2004-12/2006. Indices to proxy for these asset classes: Equities: FTSE ALL WORLD, Fixed income: Barclays Multiverse All, Real estate: FTSE EPRA/NAREIT Global, Alt. investments (proxied using commodities): S&P GSCI Commodities
Sample-based input parameters

Example 1 cont.: Change in market conditions during the financial crisis

- As of 01/2007 observe the MD for an indication of changed market conditions

![Portfolio Turbulence: MD of realized returns from historical returns (sample period 01/2004-12/2006) and 0.05/0.95 confidence bands (green)](image)

*Figure 1: Portfolio Turbulence: MD of realized returns from historical returns (sample period 01/2004-12/2006) and 0.05/0.95 confidence bands (green)*
Deviations from Model Prices

Use the MD to find deviations of observed from implied returns of theoretical/empirical models such as the CAPM or the Fama-French 3-factor model

Use...
- ...to assess how much markets are in line with models
- ...as indication for bubbles in markets
- ...to assess which competing model best describes the markets
Deviations from Model Prices

Example 2: Asset returns implied by the Black-Litterman Model [Black & Litterman, 1992]

- Use monthly returns and market capitalization of 22 country indices from 1988-2012 [Kaiser et al., 2013]
- Calculate implied returns $E(r_{t+1}) = \delta \Sigma_t w$ [Walter, 2011], assuming $\delta = 1$, weights $w_t$ and covariance matrix $\Sigma_t$ (120-month rolling window)

Figure 2: Mahalanobis distances of realized vs. Black/Litterman-implied values
Forecast Evaluation

Use of the MD to evaluate multivariate forecasts

- For a multivariate point forecast $\mu_t$, its MSE-Error matrix $\Sigma$ determines a MD-based confidence ellipsoid $\Rightarrow$ to evaluate the forecast quality regarding the observation $r_t$
- The MSE-matrix determines the shape of the confidence ellipsoid (defined by the MD) with center $\mu_t$. 
Forecast Evaluation

Two possible uses of the MD

Possible uses of the MD

1. To calculate forecast confidence regions for a (univariate) forecast path - important for many (path-dependent) financial applications [Jorda & Marcellino, 2010] (Wald Statistic based on MD)

2. Evaluate multivariate point-forecast using confidence ellipsoids [Lütkepohl, 2006] (in contrast to “Bonferroni”-confidence rectangles)
Forecast Evaluation

Example 3: Estimating a VAR-model for exchange rates

- VAR-model based on log-differences of the USD/GBP spot rate
  \( ds_t = \Delta \log(s_t) \), and the resp. forward premium \( p_t = f_t - s_t \) \( (01/1980 - 07/2013) \)
- Estimation yields the following VAR(1,1)-process:

\[
\begin{pmatrix}
    ds_t \\
p_t
\end{pmatrix} = \begin{pmatrix}
    -0.0029 \\
-0.0002
\end{pmatrix} + \begin{pmatrix}
    0.0671 & -0.8373 \\
0.0004 & 0.9025
\end{pmatrix} \begin{pmatrix}
    ds_{t-1} \\
p_{t-1}
\end{pmatrix} + \begin{pmatrix}
u_{1t} \\
u_{2t}
\end{pmatrix} \tag{4}
\]

with \( \Sigma_u = \begin{pmatrix}
    0.000888 & -0.000008 \\
-0.000008 & 0.000002
\end{pmatrix} \).
- MSE-matrices \( \Sigma_y(1) = \Sigma_u \) and

\[
\Sigma_y(2) = \Sigma_y(1) + \Phi_1 \Sigma_y(1) \Phi_1' = \begin{pmatrix}
    0.00089 & -0.00006 \\
-0.00006 & 0.00064
\end{pmatrix} \text{[Lütkepohl, 2006]}
\]
Forecast Evaluation

Example 3 cont.: Estimating a VAR-model for exchange rates

<table>
<thead>
<tr>
<th>Forecasts</th>
<th>$ds_{t+h}$</th>
<th>$p_{t+h}$</th>
<th>$ds_t(h)$</th>
<th>$p_t(h)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>06/2013</td>
<td>-0.0262</td>
<td>-0.0003</td>
<td>-0.0009</td>
<td>-0.0005</td>
</tr>
<tr>
<td>07/2013</td>
<td>0.0004</td>
<td>-0.0003</td>
<td>-0.0025</td>
<td>-0.0006</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MSE $\Sigma_y(h)$</th>
<th>2013M06</th>
<th>2013M07</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000888</td>
<td>-0.000008</td>
<td>0.00089</td>
</tr>
<tr>
<td>-0.000008</td>
<td>0.000002</td>
<td>-0.00006</td>
</tr>
</tbody>
</table>

This table reports forecasts (for 06/2013 and 07/2013) for the log return of the USD/GBP spot rate ($ds$) and the forward premium ($p$), using the VAR process stated above. In addition, it shows the realized values for these variables together with the corresponding MSE matrices.
Forecast Evaluation

Example 3 cont.: Estimating a VAR-model for exchange rates

\[ \text{log-differenced spot-rate} \]

\[ \text{forward premium} \]

\[ \text{06/2013} \]
\[ \text{07/2013} \]
\[ \text{08/2013} \]

\[ \text{Figure 3: Forecast confidence ellipses based on the Mahalanobis distance} \]
Contents

1 Introduction

2 Properties of the Mahalanobis distance

3 Types of Financial Applications

4 Conclusion
Conclusion

For multivariate (financial) problems

MD naturally supports answering questions regarding existence and magnitude of deviations

- between observations
- between observations and theoretically (model-) implied values
- between observations and predictions

implying different economic interpretations.
The end

Thank you . . .

...very much for your attention!
Conclusion

Literature I