Informeta, **L.L.C.**

Correcting Market Data Anomalies

Objectives and Conclusions

Matthew Johnson
matthew.johnson@informeta.net
Sumanth Sridhar
sumanth.sridhar@informeta.net

1 Civic Center Plaza, Suite 540
Poughkeepsie, NY 12601
www.informeta.net

Introduction

Results of risk calculations performed in Value At Risk System – VARS, depend heavily on quality of market data used in valuations. Currently, the customer's VARS used more than 200,000 different market data series that need to be cleaned and corrected. Maintenance of the market data performed by VARS Middle Office (VARS MO), and amount of market data that was reviewed was overwhelming for the personnel charged with the task.

Results of the diagnostics took more than 40 pages, and it was physically impossible to review every potential outlier reported by the diagnostics. In addition, current diagnostics produced many false positives without any indication of probability of the outlier being false or true.

The objective of the project was to use Mentys™ to evaluate existing market data and produce less outliers with better probability of the reported points being true outliers. The project was also aimed at evaluating the ability of Mentys™ to replace outliers with correct values.

There were two types of outliers – spikes and flats. Spikes are values that significantly differ from previous day's values. Flats are values that are the same as the previous day's values. Mentys™ was evaluated for its ability to recognize real spikes, real flats, and substitute outlier (spike or flat) with correct value.

Project Description

Test	Description	Steps
1	Evaluate ability of Mentys™ to identify real outlying spikes	1. Take snapshot of market data before VARS MO runs daily diagnostics. 2. Run Mentys™ on a set of market data, and obtain list of outlying spikes. 3. Run VARS MO diagnostics on the same set of data. 4. VARS MO manually identifies real outlying spikes. 5. Compare list of real spikes produced by Mentys™ on step 1.2 with the list of real spikes identified by VARS MO manually on step 1.4. Ninety percent of the two lists must be identical.
2	Evaluate ability of Mentys™ to identify real outlying flats	1. Take snapshot of market data before VARS MO runs daily diagnostics. 2. Run Mentys™ on a set of market data, and obtain list of outlying flats. 3. Run VARS MO diagnostics on the same set of data. 4. VARS MO manually identifies real outlying flats. 5. Compare list of real flats produced by Mentys™ on step 2.2 with the list of real flats identified by VARS MO manually on step 2.4. Ninety percent of the two lists must be identical.
3	Evaluate ability of Mentys™ to identify correct value for an outlier	1. VARS MO manually determines correct values for outlying spikes and flats identified on steps 1.4 and 2.4 2. Run Mentys™ on lists of outlying spikes and flats identified on steps 1.2 and 2.2, and obtain proposed values for the outliers. 3. Compare values of respective outliers from steps 3.1 and 3.2. Ninety percent of values from step 3.1 must be within five percent of values from step 3.2 for respective outliers.

Performance Summary

This section contains our analysis of the output obtained from the proof-of-concept test conducted at onsite for the Client. The figures given below are based on the list of outliers and Mentys proposals, for a single day of test data, contained in the file spreadsheet that we received from the Client.

The spreadsheet provided to us by the Client contains some entries that must be filtered out before an accurate analysis can be performed. These entries include duplicate false negatives and outliers for which the obfuscated value is unchanged from the original value. We further discount false negatives for which the difference between obfuscated and original values is negligible. There are also three instances where a Mentys proposal is noted as a "Hit," or true positive, but the obfuscated value is unchanged from the original; we reclassified these points as false positives.

Table 1. Mentys performance detecting artificially introduced outliers.

Total Outliers	True positives	False Negatives †	T.P. Rate ‡
734	665	69 †	91 % ‡

† 69 false negatives = [607 appearing in file] – [162 duplicate entries]
– [253 unchanged values] – [123 within 0.5% of original]

‡ We estimate that close consultation with experts onsite, combined with additional tuning of Mentys could increase the true positive rate.

Even when the data does not contain outliers, there remains a small, residual level of noise in the Mentys process. The false positive rate given below quantifies this independently of the frequency of introduced outliers.

Table 2. Mentys false positives.

Series analyzed †	False positives ‡	F.P. Rate
~ 366,000	1580	~ 0.43 %

† Mentys did not analyze approximately 16,000 time series due to insufficient historical data.

‡ Examination of the false positives suggests that many can be handled automatically by adding pre- or post-filters to the process, reducing the number requiring manual inspection.

The difference between the obfuscated and original values is sometimes negligible. The table below shows the varying true positive rates when different tolerance levels are used to define "negligible."

Table 3. Mentys performance by obfuscation relative difference.

Obfuscation Difference †	Qualifying Outliers	True positives	False Negatives	T.P. Rate
> 1.00 %	730	665	65	91 %
> 0.50 %	734	665	69	91 %
> 0.25 %	747	665	82	89 %
> 0.10 %	767	665	102	87 %
> 0.00 %	857	665	192	78 %

† For example, if we consider those outliers for which the obfuscated value differs from the original value by more than 0.25% tolerance, then the effective true positive rate is 89%.