Multi-Layer Outlier Detecting System (MODS)

Ronen Fluss* and Ben Z. Schreiber**

Abstract

The MODS system provides alerts of outlying developments in daily data series for the financial markets in Israel and abroad. An alert about the possibility of outlying developments in the financial markets constitutes an auxiliary tool for decision makers, including those responsible for financial stability and monetary policy. The system's uniqueness lies in the detection of outlier observations from various perspectives, and using diverse statistical procedures.

* The Information and Statistics Department, Bank of Israel

** The Information and Statistics Department, Bank of Israel, and Bar-Ilan University

Introduction

The MODS system, developed at the Information and Statistics Department, was designed to provide alerts about outlying developments in daily data series of the financial markets in Israel and abroad. An alert about the possibility of outlying developments in the financial markets constitutes an auxiliary tool for decision makers, including those responsible for financial stability and monetary policy, in assessing the state of the market and making better informed decisions.

When the system detects an outlying event (reflected by an outlier value in observed data), it sends an alert, which triggers a process of clarifying the reason for the "outlierness" (defined herein as the existence or the extremity of an outlier) and a decision to take action, if it is decided that the reason for the outlying event requires this.

The MODS monitors 50 data series daily, from four financial markets: foreign exchange markets in Israel and worldwide, interest rate markets in Israel and worldwide, capital markets, and commodities markets.

Detection of outlying developments in the system takes place on a number of levels, with each level highlighting a different aspect of outlierness, as follows:

- First layer: Each data series is monitored separately to detect outlier values in it with respect to both its long-term and short-term probability distribution.
- Second layer: Classification of a development as an outlier takes place at the level of the entire market (stocks, bonds, etc.), even if each series by itself is not an outlier (as described for the first layer)—in other words, testing the combination of the relevant market data is likely to identify an event as an outlier.
- Third layer: Classification of a development as an outlier takes place when an outstanding deviation from a statistical connection between highly correlated variables is discovered. One example of a highly correlated data series is yields on short term bonds for various maturities up to 1 year.
- Fourth layer: Classification of a development as an outlier takes place when a deviation from a (longer term) economic relation between variables is discovered. For example, the relation of the interest rate spread between the shekel and the dollar and future expectations of the shekel-dollar exchange rate (based on the Uncovered Interest rate Parity – UIP).

Detecting outliers in the various layers makes it possible to assess, from an overall perspective and from various angles, the degree of outlierness in a specific data series, or the function of a specific market from various perspectives; this capability constitutes the system's uniqueness.

The methodology of the MODS in detecting outliers in the first and second layers is described later in the review. Examples of the output of the MODS from these two layers, which are published

daily on the Information and Statistics Department's portal at the Bank of Israel, are presented. In addition, a monthly summary of the outliers in the first layer will be presented.

MODS 1 Detects Outliers—First Layer

In the first layer of the MODS, in which each series is monitored separately, a distinction is made between two types of outliers: (a) an outlier with respect to the long-term probability function, (b) an outlier with respect to the short-term probability function.

The objective: To detect outliers, both absolute and relative to their location; in other words, outliers both in terms of the long-term probability distribution and in comparison with the recent behavior of the series.

The degree of the outlier is presented in the following manner:

Red – an outlier with respect to both probability distributions

Orange – an outlier only with respect to the long-term probability distribution

Yellow – an outlier only with respect to the short-term probability distribution

Green – no outliers

The following is an example of an outlier in the "Changes in yields of 5-year unindexed government bonds" series that occurred on December 7, 2014. The following alert was sent about this outlier:

The MODS has detected an outlier in one or more of the data series on the home page of the Information and Statistics Portal. Click here for additional information.

Degree of outlierness	Last Date	Category
	05/12/2014	Foreign exchange
	07/12/2014	Inflation expectations
	07/12/2014	Yields, spreads and interest
	07/12/2014	Share indices
	05/12/2014	Commodity indices

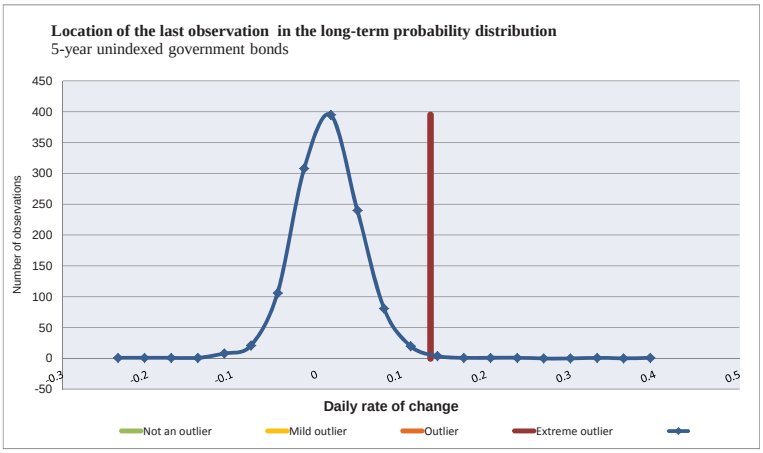
To select the series for which you will receive an e-mail alert, click here.

After clicking on "Yields, spreads, and interest," two screens are received describing two types of outliers:

1. Definition of an outlier in the long-term probability distribution

This aspect of outlierness determines whether the observation is an outlier with respect to the probability distribution based on the past five years. In order to obtain the "regular" long-term probability distribution, the most extreme observations (representing rare situations) were removed.[1]

The extreme observations are detected using the median absolute deviation (MAD)—a robust measure of the dispersion of outliers. For example, even if 30 percent of the observations are outliers, the MAD value will not be affected. In a normal probability distribution, the MAD value converges to the standard deviation. We set the default definition of an outlier as one found outside the upper 99.5 percentile and the lower 0.5 percentile of the long-term probability distribution (after removal of the extreme observations). In other words, an observation is considered an outlier if it outside the range of the 0.5 and 99.5 percentiles of the regular probability distribution. According to the example of outliers in the yield changes of government bonds, the system displays the location of the last observation with respect to the long-term probability distribution (a five-year histogram). It can be seen that the observation is an outlier from the perspective of the past five years.



[1] An extreme observation is an observation with a standard score (*Z'*) higher than 4 (absolute value). Let *X* be the values in the original series. The standard score is defined as:

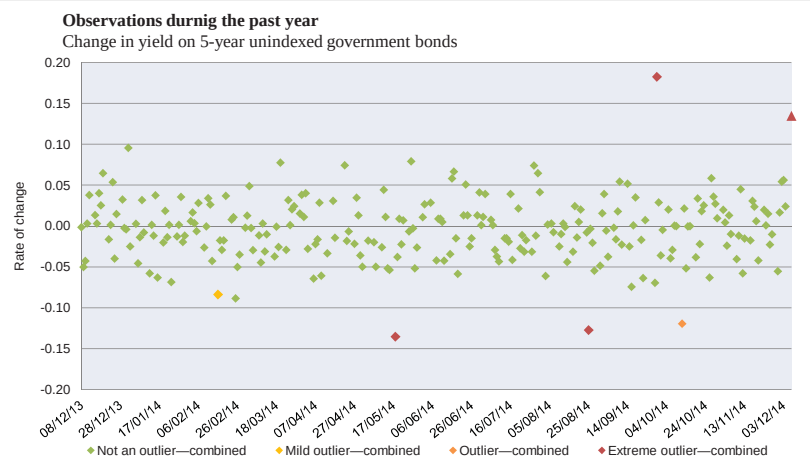
$$z' = \frac{x - Median(X)}{MAD_X}$$

$$MAD_X = 1.4286 \cdot Median\left(\left|X - Median(X)\right|\right)$$

where the medians in the formulas are calculated according to this long-term probability distribution.

2. Definition of outliers in the short-term probability distribution

This aspect of outlierness determines whether the observation is an outlier with respect to the probability distribution based on the past year. This probability distribution is defined for the standard score (*Z'*, see footnote 1), where the medians and MAD are calculated for the past 90 days for the purpose of detecting local changes in the variance. If an observation (after standardization) is found outside the upper 99.5 percentile and the lower 0.5 percentile in the short-term probability distribution (the past year) after extreme observations have been removed (*Z'*>4), it is defined as an outlier. According to the example of outliers in the yield changes of five-year unindexed government bonds, the system displays the observation of the last trading day, and all the observations during the past year.



MODS 2 Detects Outliers—Second Layer

This measure checks market outlierness by examining the joint distribution of a number of variables representing the market. Outliers are checked with respect to the long-term distribution (daily data for the past five years after eliminating trend).

The objective is to issue an alert about outliers, with respect to the long-term distribution, which indicate a possibility of irregular behavior in the market being checked, where it is sometimes possible that even if each series in itself is not an outlier, the combination of the series may be uncommon.

This part of the MODS project is a continuation of the first layer, which dealt with outliers at the univariate level. The objective in detecting outliers in multivariate data is to detect an outlier market situation on the basis of a number of market characteristics obtained in the portal with daily frequency. At this stage, the foreign exchange and stock markets are being monitored. Each market is characterized by the following characteristics: (1) rate of change, (2) trading volume, (3) implied volatility (hereafter, IV, currently relevant only to the foreign exchange and stock markets)

or intraday spread (the spread between the high and low prices). The added value of detecting an outlier market situation, in comparison with detecting univariate outliers, as in the first part of the project is as follows:

a. Multivariate analysis of the market situation makes it possible to detect situations in which each series is not an outlier in itself, but the combination is outlying.

b. The series displayed in the portal usually include yield or price changes, but do not include risk and liquidity measures, such as trading volume, the spread between the high and low intraday price, and IV.

c. The analysis makes it possible to detect special situations which do not differ from each other from a statistical perspective, but which are of great importance from an economic standpoint, particularly for policy managers. These cases include market failure, defined in this paper as a situation with low trading volumes, combined with a sharp change in prices and a high IV; and "waiting for an event," defined as a situation in which IV is high, trading volumes are very low, and price changes are negligible.

The outlier character being searched for is global, not local, meaning that its detection is being sought with respect to the long-term distribution of the series, not with respect to the short-term distribution of that period. The outlierness with respect to the local distribution was sufficient at the level of the individual series, and was already examined at the (univariate) MODS stage.[2]

Definition of Outlierness

Outlierness is determined by the distance of each observation from the center of the joint probability distribution. This distance, called the Mahalanobis Distance (MHD), is calculated as follows:

Let X be the observation on a given day, representing a vector of a number of variables representing the market. For example, in the foreign exchange market, X can be the rate of change in the shekel-dollar exchange rate, the trading volume and the IV. The distance on day t is calculated by the following formula:

$$MHD_t = (X_t - M)'S^{-1}(X_t - M)$$

Where M is the vector of averages of the variables, and S is the matrix of covariances. M and S are estimated using the observations from the past five years. The Mahalanobis Distance reflects the distance of the daily observation from the average weighted by the covariances. In order to prevent the estimates of M and S being affected by outliers, the vector of medians was used to estimate M, and a robust estimate for S that takes into account only the central group of observations was used.

[2] At this stage, it is assumed that there is no serial correlation between the characteristics selected for the various markets, a correlation requiring a slightly different treatment than the methods selected. This assumption was tested for a selection of series, and was found to be reasonable.

The notation of the distance with the robust estimators is robust MHD. An observation is defined as an extreme outlier if its distance is outside the upper 99th percentile and a moderate outlier if its distance is between the 97.5th percentile and the 99th percentile. The outliers are classified as follows:

Red—an extreme outlier (above the 99th percentile)

Orange—a moderate outlier (above the 97.5th percentile)

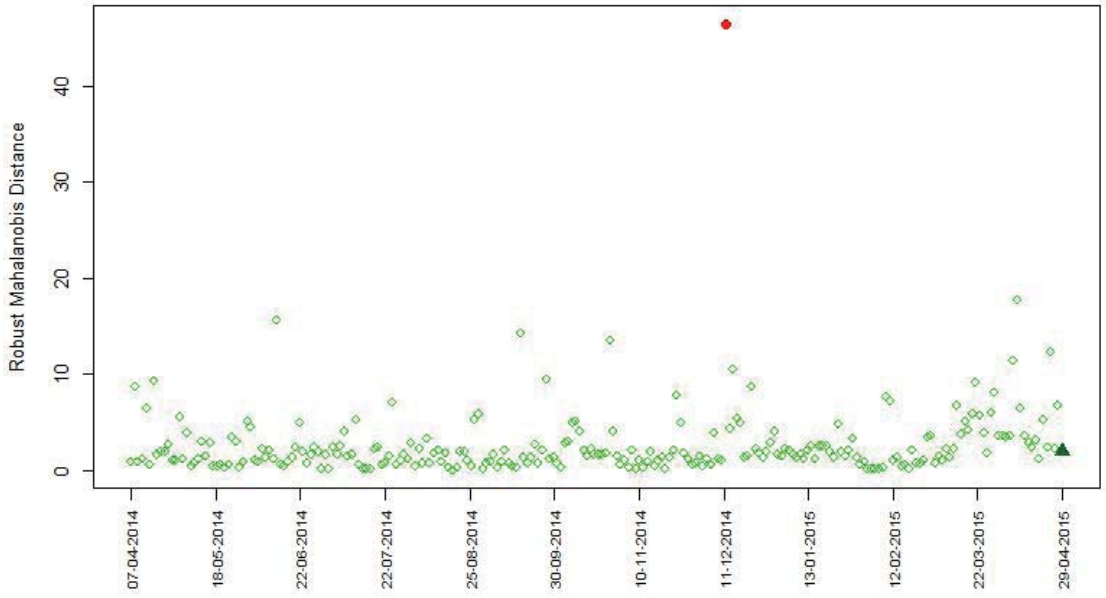
Green—not an outlier

The Graphic Representation

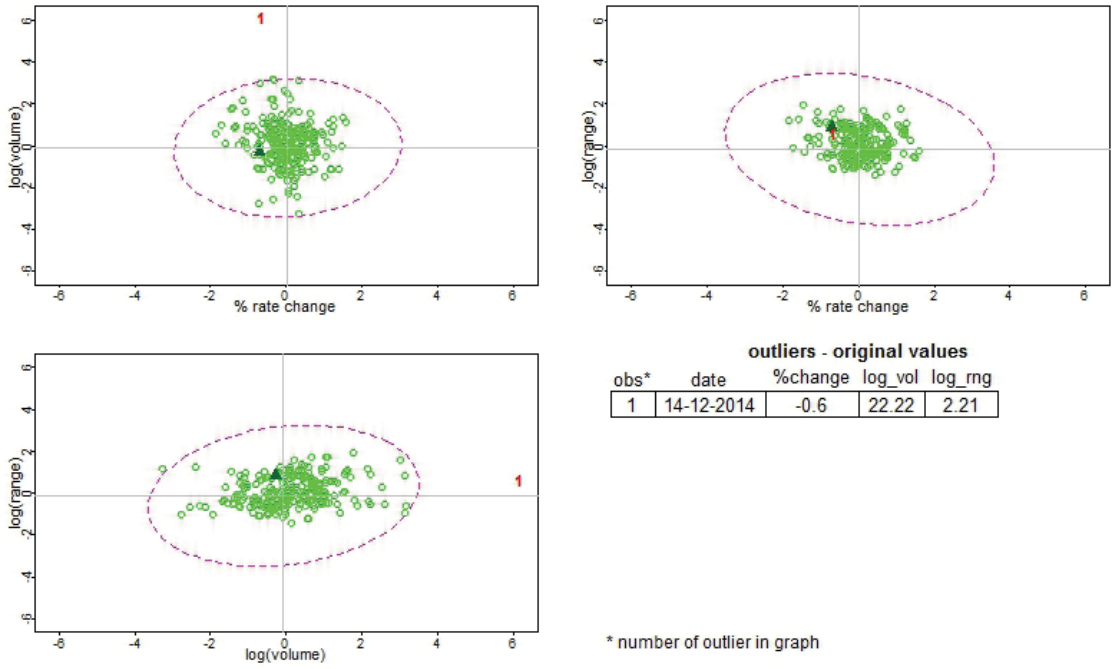
The first graph displays the robust Mahalanobis Distances of the stock market (represented by the Tel Aviv 100 Index) over the past year, with the level of outlierness noted according to the various colors (outliers are determined based on the past five years, not only the past year displayed in the graph). The distance of the current observation is marked with an enlarged triangle, whose color varies according to the market outlierness (red/orange/green).

The second graph displays the joint distribution of the three pairs of variables making up the market over the past year. This graph makes it possible to understand which of the variables have contributed to the market outliers on a given day. In each distribution, we mark the outliers, and number them according to the robust MHD (from the first graph); from the most extreme outlier to the least extreme. The original values of the outliers appear in the table below on the right. In addition, we sketch the ellipse representing the 97.5th percentile line in the bivariate normal distribution (based on the long-term distribution). This is only an approximation for the outliers limit, because a distribution of three variables, not two, is being used, and because the empirical distribution is liable to differ from the normal probability distribution. The current observation is also marked here with an enlarged triangle. The following is an example of the two graphs obtained for the stock market.

Robust Mahalanobis Distance



bivariate distribution of the variables in the last year (standardized)



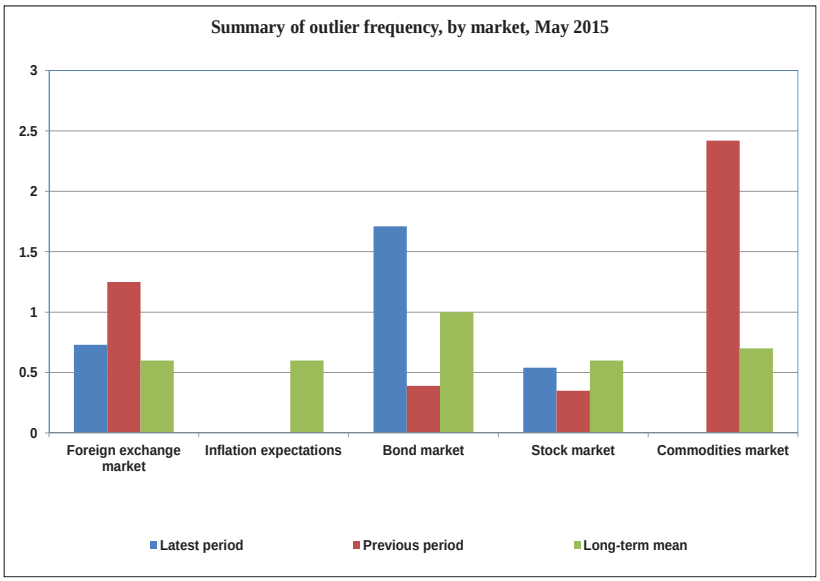
outliers - original values

obs*	date	%change	log_vol	log_rng
1	14-12-2014	-0.6	22.22	2.21

* number of outlier in graph

In the outliers table, the outliers are sorted from highest to lowest, and three variables are displayed in it (with their original values) according to which the market outlierness was determined.

One of the insights that can be deduced from the first layer in the MODS system is whether certain series or markets have been volatile recently, compared with the long-term distribution. For this purpose, the system creates a table and a graph summarizing the number of recent outliers (in the past 30 days), compared with the preceding period (the 30 days preceding them), and compared with the long-term probability distribution (the past five years—average and standard deviation) for all 50 examined series. According to these parameters, policy makers, analysts, and investors can assess whether the recent period was relatively "peaceful" or "tumultuous." The following graph displays the average numbers of outliers in a series (domestic, global, and combined outlier) in each of the five markets being examined.



As indicated by the graph, in May 2015 there were 1.71 outliers in the bond market, on average for the series, compared with a long-term average of one outlier in each series. On the other hand, no outliers at all were detected in the commodities market in May, while the number of outliers in April averaged 2.4. These figures are consistent with the high volatility in global bond prices abroad and in Israel in May, and in oil prices in April.