Robust Parameter Estimation for Financial Data Simulation

Lee, David

FinPricing

August 2025

Robust Parameter Estimation for Financial Data Simulation

David Lee

Abstract

Financial market data are known to be far from normal and replete with outliers, i.e., "dirty" data that contain errors. Data errors introduce extreme or aberrant data points that can significantly distort parameter estimation results. This paper proposes a robust estimation approach to achieve stable and accurate results. The robust estimation approach is particularly applicable for financial data that often features the three situations we are protecting against: occasional rogue values (outliers), small errors and underlying non-normality.

Key words: robust parameter estimation, financial market data, market data simulation, risk factor.

JEL Classification: E44, G21, G12, G24, G32, G33, G18, G28

1. Introduction

Simulation is widely used in financial markets for valuation and risk management. Market data that affect a financial product value and risk may change over time and impact the profit and loss of the trade. These market data called risk factors may need to be simulated for value assessment and risk management.

A simulation engine normally performs Monte Carlo simulation for all risk factors in any combination. This process will consist of successive evaluations of simulation models through time and will be generally path dependent. Those models will be fed by correlated random numbers and parameters obtained from calibration process.

Most financial valuation and risk models are based on the assumption of multivariate normality. But the real market data distributions are rarely normal. Instead, they usually have fat tails and ae contaminated by outliers. As such, the traditional estimation methods that are optimal for uncontaminated clean data have difficult to correctly be applied to actual market date due to outlier, missing points, and fat tails.

In order to achieve stable estimation of financial market data, the robust parameter estimation is deemed necessary. Robust estimation is an estimation technique which is insensitive to small departures from model assumptions, such as outliers. Robust means that changing a small part, even by a large amount, of the data does not cause a large change in the estimate.

There is a vast literature regarding robust parameter estimation. Xu et al. (2014) introduce a robust method to estimate the g- and -h distribution for risk management and stock return analysis. They use the model to obtain base distribution for outlier detection, Bassik et al. (2025) present a new approach for estimating parameters in rational ODE models from measured time series data. The approach does not suffer from non-robustness and does not require making good initial guesses.

Wang et al. (2015) propose several simple closed-form and robust estimators and study the breakdown points and asymptotic properties of the proposed estimators. Liu et al. (2021) present a novel Kalman robust smoother by introducing a specific reweighting approach to estimate the system parameters as well as the states when the nominal noise covariances are known.

Fujisawa (2013) discusses a robust parameter estimation for reducing a bias caused by outliers and uses a normalized estimating equation that is corrected to ensure that the mean of the weight is one.

Johnson et al. (2024) propose a hybrid estimation algorithm that guarantees convergence of the parameter estimate to the true value. The estimator is input-to-state stable with respect to a class of hybrid disturbances.

Liu et al. (2021) consider estimation problems involving constrained nonlinear systems with the unknown time-delays and unknown system parameters and propose a robust estimation formulation.

Guney (2025) employs the maximum Lq-likelihood estimation method that provides robust parameter estimation and further introduces the penalized Lq-likelihood method to select significant variables.

Zhu et al. (2021) propose a joint estimation and robustness optimization framework to mitigate estimation uncertainty in optimization problems by seamlessly incorporating both the parameter estimation procedure and the optimization problems.

This paper presents a robust estimation approach for financial simulation to achieve stable and accurate results. The methods are particularly applicable for financial data that often are often deteriorated by outliers, errors and underlying non-normality. It is a powerful tool for stable evaluation of statistical parameters.

We derive an estimation error formula for the case when the exact shape of the data distribution is unknown. We also propose an algorithm to find the closest correlation matrix to a given matrix when traditional correlation estimators cannot guarantee the positivity of the correlation decomposition.

The approaches are insensitive to a small number of large departures from model assumptions. The model ensures that changing a few samples in the data, even by a large amount, does not cause a large change in estimates.

The rest of this article is organized as follows: First we describe the robust estimation approaches. Second, we discuss the details of implementation. Then, we present the empirical results. Finally, the conclusions are provided.

2. Robust Estimation

Robust estimations have been used in data modelling for decades. The first robust estimator is arguably the linear regression based on combination of sample variance and the sum of absolute values in the tails. However, the estimator does not satisfy all the robustness criteria.

The sensitivity of an estimator to unbounded outliers is usually quantified using the breaking point that is defined as the fraction of points in samples whose unbounded errors do not send the total variance of the parameter estimation to infinity.

In terms of the breaking point, the most robust estimators are median absolute deviation (MAD) and interquartile distance (IQD). In this paper, we propose an integrated Median Absolute Deviation (MAD) algorithm that is a robust measure of the variability of a univariate sample of quantitative data.

The MAD for a univariate date set X_1,X_2,…,X_n is defined as

$$MAD_X = med\{[X - med_X]\} \quad (1)$$

where med_x is the median of the time series x_i

There is a constant scale factor between the standard deviation σ and the MAD, i.e.

$$\sigma = \frac{MAD_X}{K} \quad (2)$$

We extend the univariate MAD into the bivariate case. Define a bivariate MAD for two random variables X and Y as

$$MAD_{XY} = med\left\{\frac{[X-med_X][Y-med_Y]}{\sigma_X \sigma_Y}\right\} \quad (3)$$

Many financial market data have extended period of so-called sticky prices, over which the value of the time series does not change and hence the returns are zero. Another group of estimators are based on a certain way of data trimming. Data trimming methods heavily depend on the number of points trimmed, which effectively becomes the major parameter of the estimator and stipulates its breaking point.

The statistical median is an order statistic that gives the ‘middle’ value of a population sample. That is, the value such that an equal number of samples are less than and greater than the value.

The median is less sensitive to outliers than the mean, making it useful as a robust estimation technique. A robust estimator is insensitive to small perturbations from identical assumptions such as those encountered in noisy and infrequently sampled financial data.

Assume that Y is equal to X. Then we have

$$MAD_{XX} = med\left\{\frac{[X - med_X]^2}{\sigma_X^2}\right\} \quad (4)$$

Extending from the univariate MAD, we assume that there is a constant scale factor between the variance σ_X^2 and $med[(X - med_X)^2]$. The variance σ_X^2 of X can be estimated by

$$\sigma_X^2 = \frac{med[(X - med_X)^2]}{K} \quad (5)$$

The K can be calibrated by taking the error minimization as

$$min|\sigma_X^2 - mean((X - med_X)^2)| \quad (6)$$

For real time-series data, we can easily calculate MAD_{XX} and then get σ_X^2 according to (5).

Pearson's correlation is probably one of the most used statistical quantities. But it can seriously be affected by only one outlier. Its influence function is unbounded. There are several categories of the robust correlation estimations. They are median correlation measures, rank correlation measures, and Winsorized correlation measures.

Median correlation utilizes the median and a generalization of the MAD (median absolute deviation). The most popular median correlation statistics are Quadrant correlation and Percentage Bend Correlation.

Rank correlation is the study of relationships between different rankings on the same set of items. A rank correlation coefficient measures the correspondence between two rankings and assesses its significance. Two of the most popular rank correlation statistics are Kendall's tau and Spearman's rho.

Winsorizing or Winsorization is the transformation of statistics by transforming extreme values in the statistical data. A typical strategy is to set all outliers to a specified percentile of the data. Note that Winsorizing is not equivalent to trimming that simply excludes data.

In a trimmed estimator, the extreme values are discarded; in a Winsorized estimator, the extreme values are instead replaced by certain percentiles.

Extending from the univariate MAD, we also assume that there is a constant scale factor between the correlation ρ and the MAD_{XY}, that is

$$\rho = \frac{MAD_{XY}}{K} \tag{7}$$

Assume that Z_1 and Z_2 are independent t-distribution random variables (0 mean, 1 variance, and D degree of freedom). We can construct two correlated random variables X and Y with a correlation ρ such as

$$X = Z_1$$
$$Y = \rho Z_1 + \sqrt{1-\rho^2} Z_2 \tag{8}$$

We can calculate the bivariate MAD_{XY} according to (3). Therefore, we can build a relationship between the ρ and the MAD_{XY}.

When ρ =1, $K = MAD_{XY}$. Based on the known K, we can calibrate the freedom D of the t-distribution by setting $K = MAD_{XY}$.

After computing the calibrated freedom D, we can re-construct the two independent t-distribution random variables Z_1 and Z_2. The two correlated variables X and Y will be generated by recognizing the fact that the correlation increases when X and Y become bigger. The construction is given by

$$X = Z_1$$
$$Y = \tilde{\rho} Z_1 + \sqrt{1-\tilde{\rho}^2} Z_2$$

where

$$\begin{cases} \tilde{\rho} = \rho + f\left(\sqrt{X^2 + Y^2}\right) \times (1 - \rho) & if \ \rho > 0 \\ \tilde{\rho} = \rho - f\left(\sqrt{X^2 + Y^2}\right) \times (1 + \rho) & if \ \rho < 0 \end{cases} \quad (9)$$

$$f(Z) = \frac{2}{\pi} arctan(\beta Z)$$

Next, we can build a relationship in the form of a lookup table between ρ and MAD_{XY} for each β.

Based on real time-series data, we can estimate β by computing the error minimization as

$$min[mean(PearsonCor - MadCor)] \quad (10)$$

where PearsonCor is the Pearson correlations and MadCor is the correlation estimated by searching the lookup table. After we know the β, we can finally create the lookup table between the ρ and the MAD_{XY}.

The distribution is estimated as a t-distribution with degrees of freedom parameter ν= 4.5. The value of ν was selected by calculating the estimation error in recovering a known volatility from simulated data from a t-distribution. The estimation error was shown to exhibit a small spread across all input distributions with $\nu > 2$ for values of the estimation parameter $4 \leq \nu \leq 5$.

Consider a stochastic representation of a t-distributed random variable,

$$X_t = \mu + \sigma \frac{Z_t}{S_t}; t = 1,2,...,n$$

where Z_t are independent standard normal variables that are also independent of S_t, and S_t are independent $\chi^2_\nu/(\nu\text{-}2)$ random variables. We can determine an iterative scheme for calculating the mean and variance of the distribution. The following set of equations is solved iteratively, converging to σ-hat, μ-hat for sufficiently large k,

$$\hat{\sigma}_{k+1}^2 = \frac{1}{n}\sum_{t=1}^{n} w_t^k (X_t - \hat{\mu}_k)^2$$

$$\hat{\mu}_{k+1} = \frac{\sum_{t=1}^{n} w_t^k X_t}{\sum_{t=1}^{n} w_t^k}$$

$$w_t^{k+1} = \frac{\nu+1}{\nu-2}\left(1 + \frac{(X_t - \hat{\mu}_{k+1})^2}{(\nu-2)\hat{\sigma}_{k+1}^2}\right)^{-1}$$

where ν is the degree of freedom and is set to 4.5

The original estimate does not affect the final result of the iterations, but only the number of iterations, i.e., rate of convergence. The current version uses the sample median and sample standard deviation as the starting point for the mean and volatility respectively. The iterations are repeated until the following convergence criterion is met:

$$\left|\frac{\hat{\sigma}_{k+1}^2 - \hat{\sigma}_k^2}{\hat{\sigma}_k^2}\right| \le 1e^{-5}$$

3. Implementation Procedure

The calibration procedure consists of several steps: First, the calibration of K in equation (5) is based on real time-series data. For risk factor (time-series) X, we calculate the MAD_{XX} according to (4) and the $mean((X - Med_X)^2)$. The variance V can be estimated following (5). Then the K can be estimated by taking the error minimization (6).

Next, the calibration of freedom D of the t-distribution is based on Monte-Carlo simulation. We construct two independent t-distribution random variables Z_1 ~ t(D, 0, 1) and Z_2 ~t(D, 0, 1). Furthermore, we can set two correlated random variables X and Y with correlation ρ according to (8).

By changing ρ and then calculating MAD_{XY}, we can build the relationship between ρ and MAD_{XY}. There are different relationship lookup tables for different freedoms. Then the freedom of the t-distribution can be estimated by setting $K = MAD_{XY}$ in the case of ρ = 1.

Then the calibration of beta β combines Monte-Carlo simulation and real time-series data. Based on the freedom of the t-distribution determined above, we re-generate Z_1 and Z_2. Then we can construct X and Y according to (9). If β is known, we can build the relationship lookup table between the ρ and the MAD_{XY}.

For real time-series data, we can get robust correlations by looking up the table and also calculate Pearson correlations. The β can be estimated by taking the error minimization (10).

Finally, we can create a relationship lookup table between the ρ and the MAD_{XY} based on β.

4. Numerical Results

To verify the assumption that there is a constant scale factor between the correlation ρ and the MAD_{XY}, we conduct some convergence tests. If the MAD_{XY} converges to a constant, we believe that we prove the assumption empirically. The convergence results are shown in Table 1. The Figure 1 shows the case where freedom=4.

Table 1 Convergence results for ρ = 0.5 (n ~ freedom)

N	MAD				
	n=3	n=3.5	n=4	n=4.5	n=5
100	0.079727	0.096331	0.109184	0.117092	0.123175
200	0.079664	0.09697	0.108736	0.116778	0.122957
300	0.079735	0.096989	0.108652	0.116994	0.123148
400	0.079747	0.097028	0.108591	0.117023	0.123103

500	0.079686	0.096991	0.108645	0.116964	0.123127
600	0.079682	0.096998	0.108675	0.116984	0.123137
700	0.07967	0.097024	0.108669	0.116958	0.123129
800	0.079669	0.097025	0.108655	0.116979	0.123146
900	0.07966	0.097025	0.108671	0.11694	0.123136
1000	0.079674	0.097017	0.108646	0.116932	0.123139
1100	0.079693	0.097039	0.108657	0.116958	0.123125
1200	0.079702	0.097035	0.108656	0.116968	0.123136
1300	0.079685	0.097025	0.10866	0.116952	0.123129
1400	0.07968	0.097024	0.108665	0.116955	0.12314
1500	0.079686	0.097025	0.108663	0.11695	0.123125

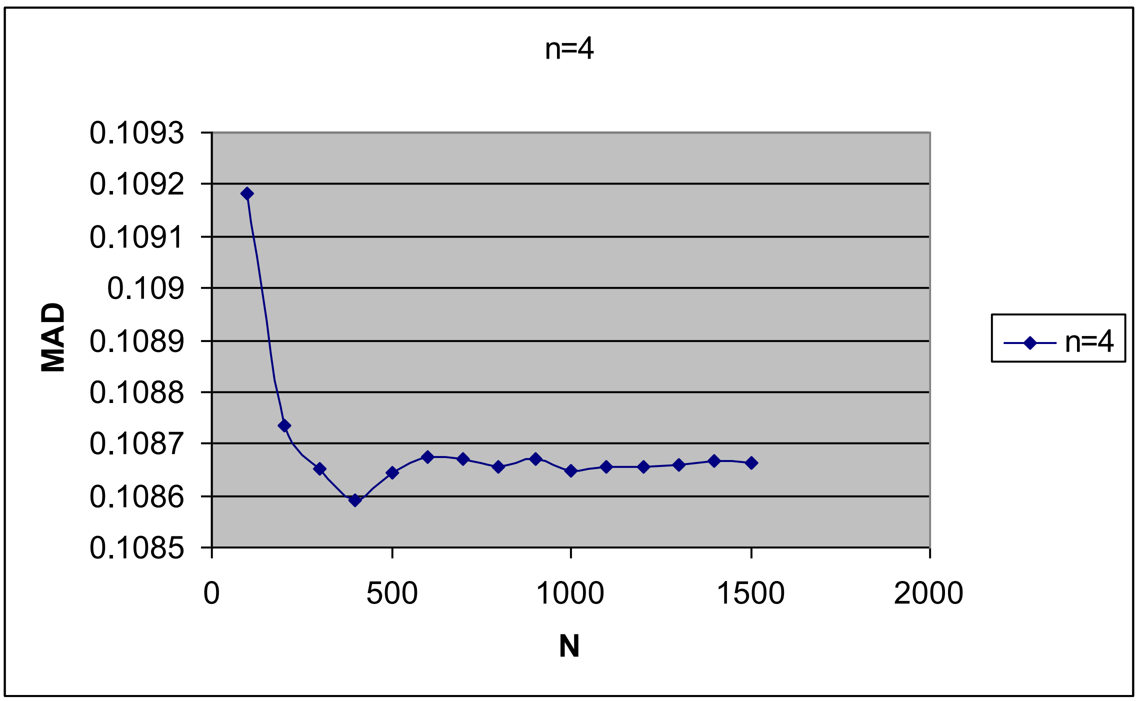

Figure 1: Convergence results of MAD correlation

(N ~ number of grids, n ~ number of freedom)

The table and figure tell us that the estimator is resistant to outliers. It is not sensitive to the distribution assumption and at the same time not very sensitive to the number of degrees of freedom.

The relationship lookup table between the ρ and the MAD_{XY} according to (8) is shown in Table 2. Figure 2 shows the case where freedom=4.

Table 2: MAD vs Correlation (n ~ freedom)

rho	MAD				
	n=3	n=3.5	n=4	n=4.5	n=5
-1	-0.19502	-0.24168	-0.27432	-0.29839	-0.31685
-0.95	-0.18039	-0.22338	-0.25343	-0.27558	-0.29255
-0.9	-0.1675	-0.20707	-0.23466	-0.25495	-0.27047
-0.85	-0.15529	-0.19171	-0.21693	-0.23548	-0.24962
-0.8	-0.1436	-0.17695	-0.19998	-0.21684	-0.22968
-0.75	-0.13226	-0.16271	-0.1836	-0.19887	-0.21045
-0.7	-0.12123	-0.14882	-0.16778	-0.1815	-0.19185
-0.65	-0.11048	-0.13536	-0.15235	-0.16466	-0.17388
-0.6	-0.09998	-0.12225	-0.13739	-0.14828	-0.15641
-0.55	-0.08971	-0.10947	-0.12283	-0.13234	-0.13953
-0.5	-0.07967	-0.09702	-0.10865	-0.11693	-0.12314
-0.45	-0.0699	-0.08492	-0.09491	-0.10201	-0.10729
-0.4	-0.06038	-0.07316	-0.08163	-0.08761	-0.092
-0.35	-0.05114	-0.0618	-0.0688	-0.07374	-0.07735
-0.3	-0.0422	-0.05085	-0.05649	-0.06044	-0.06333
-0.25	-0.03363	-0.04037	-0.04476	-0.04779	-0.05003
-0.2	-0.02545	-0.03045	-0.03368	-0.03591	-0.03755
-0.15	-0.01779	-0.0212	-0.0234	-0.02492	-0.02602
-0.1	-0.01075	-0.01277	-0.01407	-0.01496	-0.01562
-0.05	-0.0046	-0.00545	-0.006	-0.00638	-0.00666

0	-3.44E-20	-4.30E-20	-4.90E-20	-5.40E-20	-5.80E-20
0.05	0.004596	0.005452	0.005998	0.006384	0.006661
0.1	0.010754	0.012769	0.014066	0.01496	0.01562
0.15	0.017786	0.021196	0.023396	0.024922	0.026021
0.2	0.025449	0.030451	0.033676	0.035914	0.037551
0.25	0.033626	0.040371	0.044765	0.047795	0.05003
0.3	0.042203	0.050851	0.056488	0.06044	0.063326
0.35	0.051142	0.061801	0.068799	0.073743	0.077352
0.4	0.060379	0.073157	0.081628	0.087614	0.091996
0.45	0.069905	0.084922	0.094915	0.10201	0.107289
0.5	0.079674	0.097017	0.108646	0.116932	0.123139
0.55	0.089708	0.109473	0.12283	0.13234	0.13953
0.6	0.099984	0.122248	0.137393	0.148277	0.156406
0.65	0.110485	0.135362	0.152346	0.164655	0.173881
0.7	0.121229	0.148824	0.167781	0.181504	0.191846
0.75	0.132265	0.162706	0.183605	0.19887	0.210453
0.8	0.143596	0.176947	0.199984	0.216842	0.22968
0.85	0.15529	0.191706	0.216934	0.23548	0.24962
0.9	0.167501	0.207073	0.234663	0.254945	0.270472
0.95	0.18039	0.223377	0.253433	0.275585	0.292552
1	0.195022	0.241677	0.274319	0.298388	0.316847

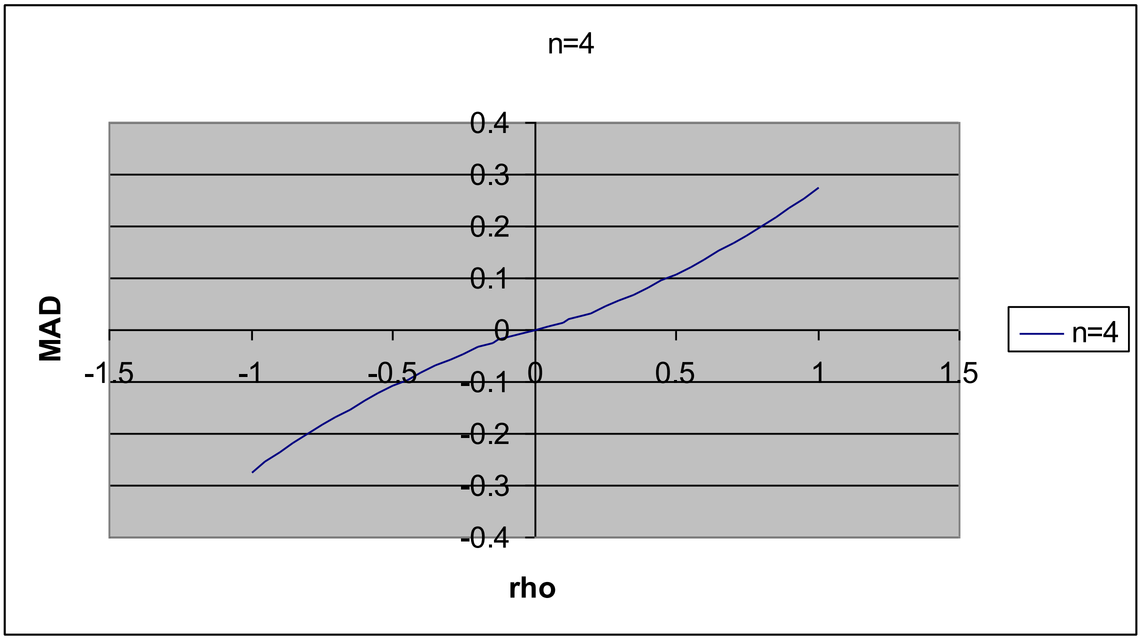

Figure 2 Relationship between correlation and MAD
(t distribution: freedom n=4)

The above table and figure show the error relative to the estimated correlation for different degrees of freedom. The relationship between correlation and MAD is also plotted. They also show that the estimators with the degrees of freedom between 4.5 and 5 are practically indistinguishable. At the same time, all estimators shown above are relatively insensitive to the data contamination.

Consider an example of financial time series drawn from the commodity market. The data source consists of natural gas NYMEX futures and forwards on a pipeline from Jan 2015 to Sept 2022. The estimated parameters for commodity are: K = 0.33, Freedom D = 5.43 and β = 0.1455. The estimated errors are:

Table 3: estimated errors for commodity market data

Mean	STD	Max	Min	SSE
1.5864e-005	0.1002	0.7434	-0.5866	218.2184

The final relationship between the ρ and the MAD_{XY} is shown in Figure 3.

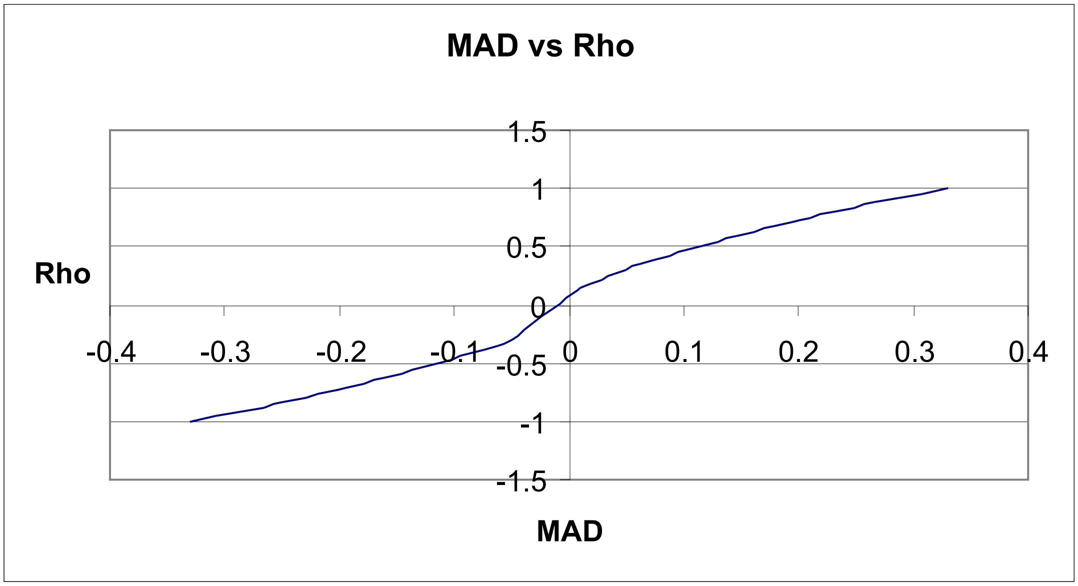

Figure 3: ρ vs MAD_{XY} for commodity

The volatility correction is derived from the implied volatility time series. The average implied volatility is calculated for each pipeline (NG, AECO, CHCG, ELSAN, NWROC, PAN, SOCAL) by simply taking the average of the 9-year time series. Then, the average implied volatility for a given calendar month is calculated by averaging implied volatilities in the time series that fall within that month. This is repeated for each calendar month, resulting in 12 calendar month average volatilities for each of the 7 pipelines listed above. The volatility correction is then given by the following ratio:

$$C = \frac{\overline{\sigma}_{month}}{\overline{\sigma}_{average}}.$$

We also use real interest rate data to test the model. There are essentially two types of IR curves provided in the feeds, so called 'dependent' or 'basis' curves and 'derived' curves. The rates for 'dependent' curves are obtained by adding the spreads of 'dependent' curves on top of their corresponding 'base' curves. However, the rates supplied in the feeds for the 'dependent' curves are already 'base + spread' values, and therefore, there is no need for conversion.

Interest rates are calculated under certain day count and compounding rules. In a case ACT/ACT day count rule is used, the appropriate way to calculate particular discount factor or rate is to calculate actual # of days for each year covered by the period of interest. For example, 2016 will be a leap year, so if today is 15 Dec 2015, and the period of interest t ends on 20 Jan 2016, the appropriate t/d calculation would be $t/d = 16/365 + 20/366$. However, since every fourth year is a leap year, a common way to simplify the calculations in this case is to use a rule of ACT/365.25, i.e. assume that every year has a fixed # of days of 365.25. The last tab of the attached shows this is a very good approximation, and can be used till time allows for proper implementation of ACT/ACT rule.

The assumption is that the rates of all IR curves should be converted to their continuously compounded equivalents, while keeping the same day count rule. The idea of converting and saving the rates to follow both continuous compounding and ACT/365 day count rule is to be considered. Namely, by doing so, we might increase computational efficiency as the pricing formulas will be assuming the input rates always respect the same conventions.

The data source consists of interest rate curves from Jan 2015 to Sept 2022.The estimated parameters for interest rates are: K = 0.1, Freedom = 2.4, and β = 0.0912

Table 4: estimated errors for interest rate market data

Mean	STD	Max	Min	SSE
-5.1729e-005	0.1166	0.8892	-0.4685	32.8291

The final relationship between the ρ and the MAD_{XY} is shown in Figure 4.

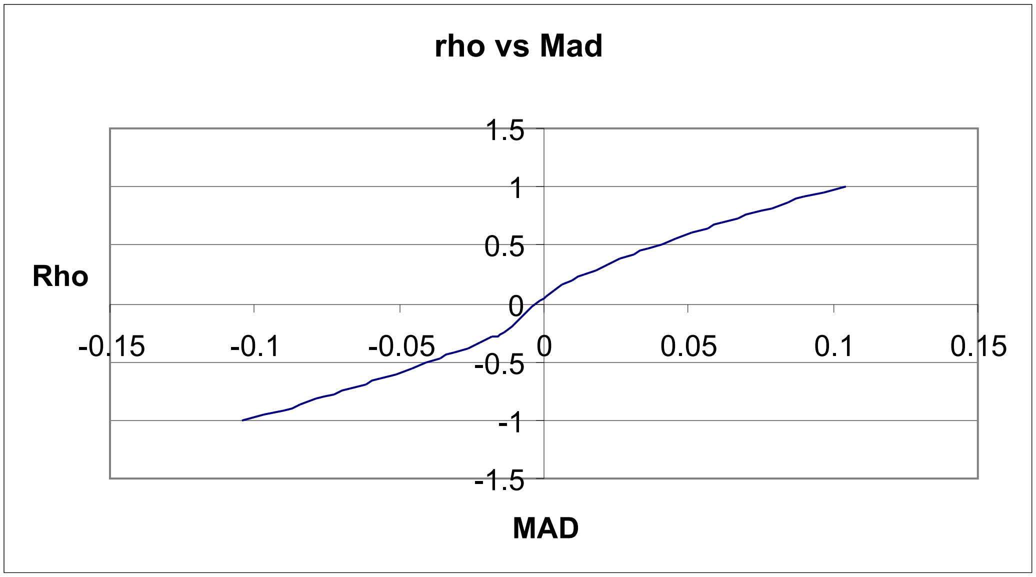

Figure 4: ρ vs MAD_{XY} for interest rates

The equity market data of interest can be subdivided into price-related data, dividends, rebates, and volatility surfaces. Note that when price-related data is in question, the market data are the last prices.

Equity prices are classified by the security ticker, exchange code and currency. The market prices of individual stocks and indices and their dividend information are required to price equity derivatives. The dividend can be reported as cash dividend or dividend yield (see https://finpricing.com/lib/EqBarrier.html).

The data source consists of equity last prices from Jan 2015 to Sept 2022. For each equity name, the integrity of the return series is checked against some criteria. The max allowed missing sets the maximum allowable ratio of values within a return series that fail to meet the criteria.

The estimated parameters for equity are: K = 0.271, Freedom = 3.94, and β = 0.1374.

Table 4: estimated errors for equity market data

Mean	STD	Max	Min	SSE
8.2619e-005	0.0856	0.6319	-0.3840	76.4512

The final relationship between the ρ and the MAD_{XY} is shown in Figure 5.

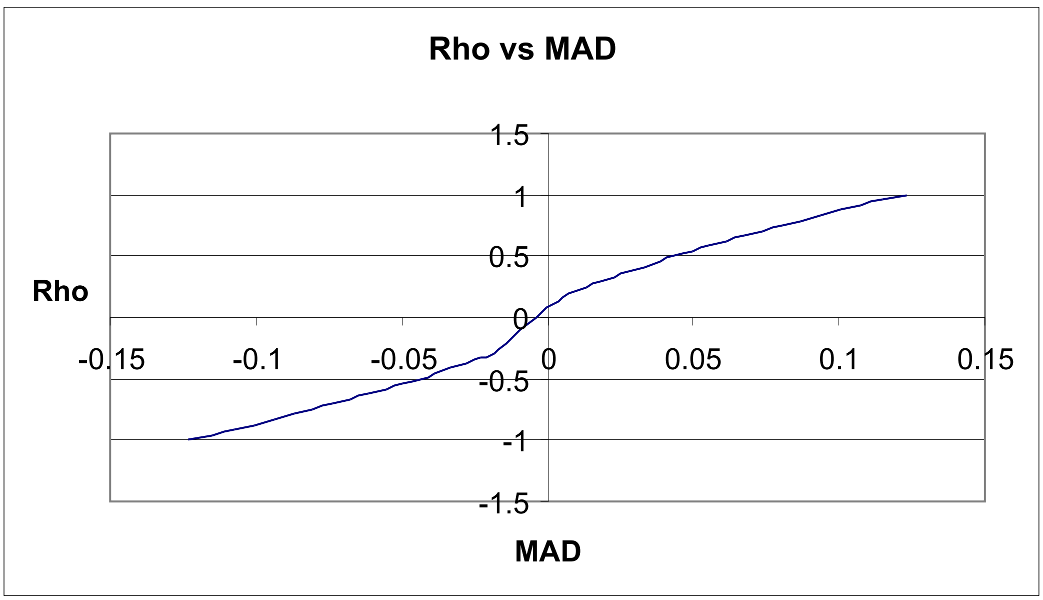

Figure 5: ρ vs MAD_{XY} for interest rates

We show the robustness comparison results below. It can be seen that the MAD approach has the smallest sum of squired changes (SSC) that means it has the best robustness.

Table 2: Robustness Comparison (Outliers occur, SSC = Sum of Squired Changes)

	MAD		Pearson		Kendal Tau		median correlation	
ρ_{XY}	value	change	value	change	value	change	value	change
1	0.45530	0	1	0	0.99988	0	1.00000	0
0.9	0.39483	0.00872	1	0.10000	0.89461	-0.00574	0.89995	0.00059
0.8	0.33001	0.00755	1	0.20000	0.79707	-0.00323	0.79912	-0.00061
0.7	0.27181	0.00643	0.999999	0.30000	0.69999	0.00006	0.70004	-0.00068
0.6	0.21974	0.00762	0.999999	0.40000	0.60388	0.00428	0.60225	0.00508

0.5	0.17070	0.00740	0.999999	0.50000	0.50814	0.00876	0.49887	-0.00102
0.4	0.12600	0.00687	0.999997	0.60000	0.41261	0.01333	0.39847	-0.00212
0.3	0.08690	0.00615	0.99999	0.69999	0.31723	0.01790	0.30193	0.00342
0.2	0.05224	0.00451	0.99999	0.79999	0.22206	0.02237	0.20234	0.00322
0.1	0.02359	0.00394	0.99995	0.89995	0.12707	0.02667	0.09990	0.00129
0	0	0	0	0	0	0	0	0
-0.1	-0.02359	-0.00406	-0.99995	-0.89995	-0.12707	-0.02667	-0.09990	-0.00129
-0.2	-0.05224	-0.00504	-0.99999	-0.79999	-0.22206	-0.02237	-0.20234	-0.00322
-0.3	-0.08690	-0.00622	-0.99999	-0.69999	-0.31723	-0.01790	-0.30193	-0.00342
-0.4	-0.12572	-0.00663	-1	-0.60000	-0.41261	-0.01333	-0.39847	0.00212
-0.5	-0.17069	-0.00750	-1	-0.50000	-0.50814	-0.00876	-0.49887	0.00102
-0.6	-0.21974	-0.00767	-1	-0.40000	-0.60388	-0.00428	-0.60225	-0.00508
-0.7	-0.27181	-0.00660	-1	-0.30000	-0.69999	-0.00006	-0.70004	0.00068
-0.8	-0.33001	-0.00783	-1	-0.20000	-0.79707	0.00323	-0.79912	0.00061
-0.9	-0.39483	-0.00892	-1	-0.10000	-0.89461	0.00574	-0.89995	-0.00059
-1	-0.45530	0	-1	4.218E-14	-0.99988	0	-1.00000	0
SSC		0.000832		5.69976		0.00370		0.00111

5. Conclusion

The most important statistical performance criteria are robustness, resistance and efficiency. In the statistical context resistance refers to the degree of tolerance of a statistical technique to the presence of outliers. Efficiency is a relative measure of sampling variability. It relates some technique of interest to some standard/traditional technique. Robustness refers to insensitivity with regard to an underlying assumed probability model.

This paper proposes an integrated parameter estimation approach for both variance and correlation that are using in financial simulation. The approach is a robust measure of the variability of a univariate sample of financial market data.

The approach is very well fit for calculating volatilities and correlations for financial time series. It is resistant to outliers, not sensitive to the distribution assumption, and at the same time not very sensitive to the number of the degrees of freedom.

References:

Bassik, O. et al., 2025, “Robust parameter estimation for rational ordinary different equations,” Applied Mathematics and Computation, 509 (1).

Fujisawa, H., 2013, “Normalized estimating equation for robust parameter estimation,” Electron. J. Statist., 7, 1587-1606.

Guney, Y. and Arslan, O., 2025, “Robust parameter estimation and variable selection in regression models for asymmetric heteroscedastic data,” Journal of Applied Statistics, 1-38.

Johnson, R. Cairano, S. and Sanfelice, R., 2024, “Robust parameter estimation for hybrid dynamical systems with linear parametric uncertainty,” Automatica, 167, 111766.

Liu, C., Han, M., Gong, Z. and Teo, K., 2021, “Robust parameter estimation for constrained time-delay systems with inexact measurements,” Journal of Industrial & Management Optimization, 17 (1), 317-337.

Liu, Y., Wang, H. and Zhang W., 2021, “Robust parameter estimation with outlier-contaminated correlated measurements and applications to aerodynamic coefficient identification,” Aerospace Science and Technology, 118, 106995

Xu, Y., Iglewicz, B. and Chervoneva, I., 2014, “Robust estimation of the parameters of g-and -h distributions, with applications to outlier detection,” Computational Statistics & Data Analysis, 75, 66-80

Wang, M., Park, C. and Sun X., 2015, “Simple robust parameter estimation for the Birnbaum-Saunders distribution,’ Journal of Statistical Distributions and Applications, 14

Zhu, T., Xie, J. and Sim, M., 2021, “Joint estimation and robustness optimization,” 68 (3).