INFFC Data Analysis: Lower Bounds and Testbed Design Recommendations

Radu Drossu and Zoran Obradović

School of Electrical Engineering and Computer Science
Washington State University, Pullman, Washington, 99164-2752

Abstract

This article summarizes the results of an exploratory data analysis on the INFFC competition time series. The analysis provides evidence that the problem is non-stationary and that the interpolation process for filling-in missing values alters the data distribution. The accuracy for trivial and linear predictors, determined in order to establish accuracy lower bounds for reasonable nonlinear prediction systems, identifies competition entries with prediction accuracies below the provided bounds. Finally, testbed design recommendations for future financial time series competitions are extracted from the results of this analysis.

1 Introduction

The objective of the First International Nonlinear Financial Forecasting Competition (INFFC) was to evaluate financial time series forecasting models on a pre-specified benchmark problem. The INFFC benchmark data [7] was a cotton futures intra-day time series comprising 107,386 non-uniformly sampled 6-tuples consisting of time stamp, opening, highest, lowest, and last strike price of the minute, along with the tick volume (the number of strike prices collected in the one minute period). The first 80,000 samples were provided to the competitors for model design and verification, whereas the last 27,386 samples were used by the INFFC panel for evaluating submitted forecasting systems. In order to provide forecasts for prediction horizons of 120 minutes and 1 day ahead, an interpolation process had to be performed in which the missing price values were obtained by repeating the last available closing price. This resulted in approximately 261,000 training samples and 67,000 test samples.

The objectives of this article are to analyze the competition time series, to establish accuracy lower bounds for reasonable nonlinear prediction systems, and to provide testbed design recommendations for future financial forecasting competitions (a more detailed report can be found in [4]). It is important to emphasize that the predictors discussed in this article are not designed as competition entries, and no attempt is made whatsoever to evaluate the adequacy of the submitted forecasting systems.

2 Statistical Analysis of the Competition Data

The INFFC call for participation did not explicitly specify whether the goal of the competition was to predict the closing price or the price change, and consequently this article considers both objectives. An analysis is performed in order to determine: (1) whether the interpolation process explained in the Introduction alters the data distribution; (2) whether the distribution of the data set provided to the competitors (*training set*) is the same as the distribution of the data set used by the INFFC panel to evaluate prediction accuracy (*test set*); (3) whether a reliable prediction horizon can be estimated from autocorrelation plots. In addition to a visual inspection of normalized histograms (the number of points in each bin is divided by the total number of points) and autocorrelation plots, the analysis also includes the chi-square and Kolmogorov-Smirnov tests for comparing whether two data distributions are different [6].

2.1 Interpolation Effects on Data Distribution

Normalized histograms for un-interpolated and interpolated training and test set prices suggested different distributions (histograms for un-interpolated and interpolated price test data are shown in Figs. 1 and 2). More objectively, both the chi-square and the Kolmogorov-Smirnov tests yielded a probability 0.999 of

rejecting the null hypothesis that the un-interpolated and the interpolated price data distributions are the same, both for the training and the test sequences.

A similar analysis on price changes indicated the histograms of the interpolated data sets to be significantly more leptokurtic (pointed) than the corresponding histograms of the un-interpolated data sets (see Figs. 3 and 4). This is due to the fact that a fairly large amount of data was introduced through the interpolation process by repeating the last available closing price when data was missing, resulting in a large number of zero-valued price changes shown as a long bar centered about the zero value of the interpolated histograms. This visual finding that the interpolation alters the price change distribution is confirmed by chi-square and Kolmogorov-Smirnov tests that both rejected the null hypothesis with probabilities between 0.93 and 0.999.

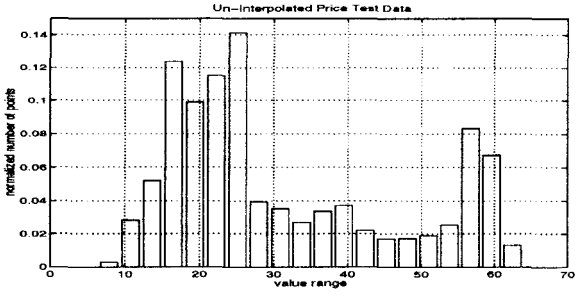

Figure 1: Un-interpolated Price Test Set

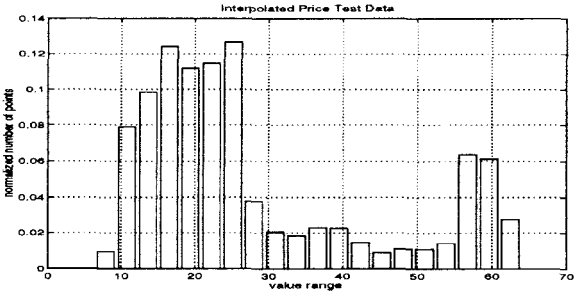

Figure 2: Interpolated Price Test Set

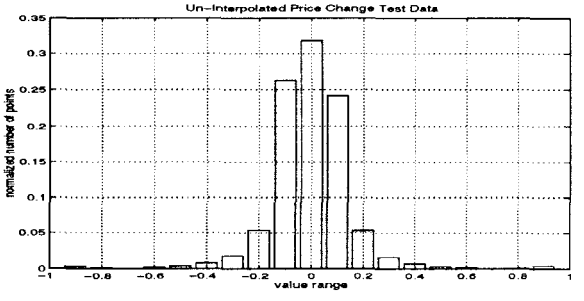

Figure 3: Un-interpolated Price Change Test Set

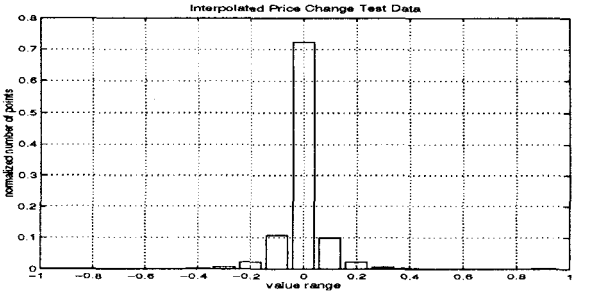

Figure 4: Interpolated Price Change Test Set

2.2 Stationarity and Data Correlation Analysis

A *non-stationary* time series can be described as a time series whose characteristic parameters change over time [5]. In general, non-stationarity detection can be reduced to identifying two sufficiently long, distinct data segments that have significantly different statistics (distributions).

For the competition data, stationarity analysis was reduced to comparing the interpolated training and test distributions for the price, as well as for the price change time series. Both for the price and the price change data, the histograms indicated significantly different training and test distributions, confirmed also by the chi-square and Kolmogorov-Smirnov tests that rejected the null hypothesis, according to which the two distributions were the same, with probability 0.999.

Additional tests on un-interpolated price and price change time series were performed in order to determine whether the non-stationarity was an intrinsic property of the original cotton time series or it has been artificially introduced by the interpolation process. The results confirmed that the original time series also exhibited non-stationarity.

Autocorrelation plots for a 300 samples data segment from the training and the test parts of the interpolated price time series exhibited a relatively slow drop, thus suggesting a potentially large reliable prediction horizon when predicting actual prices. Similar autocorrelation plots for the interpolated price change time series suggested a very short reliable horizon for price change prediction using previous price changes only.

Predictor	*DS*	*modDS*	*nRMSE*
Random Walk	7.888	52.295	0.075
Mean	0.000	64.255	1.161
AR(3)	13.231	40.949	0.265
ARI(1)	7.899	50.787	0.267

Table 1: Price 120 Minutes ahead

Predictor	*DS*	*modDS*	*nRMSE*
Random Walk	16.745	48.060	1.421
Mean	0.000	49.397	1.000
AR(3)	21.944	44.161	1.318
ARI(1)	17.601	47.036	1.402

Table 2: Price Change 120 Minutes ahead

3 Trivial and Linear Predictors

Assuming large training data sets and sufficient training time, it is reasonable to expect non-linear forecasting systems to perform at least as well as trivial or linear predictors. Hence, it is important to determine the prediction accuracy of both trivial and linear predictors in order to establish lower bounds for the prediction accuracy of reasonable non-linear predictors. The trivial predictors considered in this analysis were the random walk and the mean predictors. The random walk predictor considers a future prediction to be equal to the last available process value, whereas the mean predictor generates future predictions as being equal to the mean of the training data samples. The linear forecasting [1] considered in this article is based on autoregressive predictors of order p, denoted as AR(p), in which the predicted process value at time t, $\hat{x}_t$, is obtained as a linear combination of p previous process values, $x_{t-1}, \ldots, x_{t-p}$. The analysis also considered the autoregressive-integrated ARI(p) model which is an AR(p) model applied to differenced data.

The accuracy measures for comparing the actual data sequence $\{x_t\}$ and the predicted data sequence $\{\hat{x}_t\}$, reported in this article are the normalized root mean squared error ($nRMSE$) and a modified directional symmetry ($modDS$) defined as

$$modDS = \frac{100}{n}\sum_{t=1}^{n} c_t; \quad c_t = \begin{cases} 1, & \text{if } (x_t - x_{t-1})(\hat{x}_t - \hat{x}_{t-1}) > 0 \text{ and } |x_t - x_{t-1}| > \epsilon \text{ and } |\hat{x}_t - \hat{x}_{t-1}| > \epsilon \\ & \text{or} \\ & |x_t - x_{t-1}| < \epsilon \text{ and } |\hat{x}_t - \hat{x}_{t-1}| < \epsilon \\ 0, & \text{otherwise,} \end{cases}$$

ϵ being a small constant related to the numerical precision involved in the $modDS$ computation.

The $nRMSE$ measure is always non-negative with smaller values indicating a better predictor. The standard directional symmetry DS measures the percentage of correctly predicted market directions, with larger values suggesting a better predictor [2]. Unfortunately, the DS is meaningless for a time series with a large number of equal consecutive value pairs, as is the case for the interpolated competition time series, since it accounts only for correctly predicted upward and downward trends. Consequently, the proposed $modDS$ takes into consideration all the correctly predicted directions (upward, downward, and no change), as well as computer truncation errors.

The $nRMSE$ as a function of prediction horizon for a trivial random walk predictor on the price test series exhibited a relatively slow rise, suggesting a fairly easy prediction problem both for 120 minutes and 1 day prediction horizons, the $nRMSE$ values for 120 minutes ahead being shown in Table 1. The random walk predictor yielded the best $nRMSE$ value both compared to the mean predictor and to the linear autoregressive models. The same was true also for the 1 day ahead prediction. The investigated AR(p) and ARI(p) models based on a sampling rate of one minute were restricted to orders p$\leq$ 10. Consequently, these linear models did not include the random walk predictor as a particular case for the competition's prediction horizons (e.g. for 120 minutes horizon, the AR(p) model has to be at least of order p= 120). Although the $nRMSE$ value for the mean predictor was extremely poor, its $modDS$ value for price prediction was significantly better than those of the other predictors. However, this is not a particular achievement of the mean predictor, but an artifact of the interpolation process, since $|\hat{x}_t - \hat{x}_{t-1}| < \epsilon$ is always true for the mean predictor, whereas $|x_t - x_{t-1}| < \epsilon$ is true for each t introduced by the interpolation process.

The price change prediction results for a 120 minutes horizon are presented in Table 2. As expected, the $nRMSE$ value for the random walk predictor was significantly larger as compared to that obtained on the price prediction, confirming the increased difficulty of the problem. However, the $nRMSE$ value for the mean predictor improved for the price change prediction since the mean of the price change training series was the same as the mean of the price change test series, while this was not true for the means of the

actual prices. It was also evident that the most appropriate AR(p) model (p= 3) performed slightly better than the random walk predictor with respect to the $nRMSE$ measure. The decrease in $modDS$ for the mean predictor was due to the fact that for the price change prediction, c_t is a function of three consecutive price values instead of two for the price prediction, thus reducing the number of cases in which the actual price change trend coincides with the predicted price change trend. For price change prediction, the mean predictor appeared to be better than the random walk and the AR predictors, with respect to both $nRMSE$ and $modDS$. Similar observations were valid also for the 1 day ahead price change prediction.

It is important to observe that the lower bounds for reasonable nonlinear predictors obtained through this simple analysis are probably weak. Better lower bounds can be obtained by investigating AR models of higher order (e.g. when predicting 120 minutes ahead one might want to use information from at least the previous 120 minutes) and considering different data sampling rates [3]. However, this was not necessary for this analysis, since even these simple predictors were able to outperform some of the competition entries [7].

4 Conclusions and Recommendations for Future Competitions

This article reported the results of an exploratory data analysis, as well as the prediction accuracy of random walk, mean and autoregressive predictors on the INFFC competition time series.

The exploratory data analysis should be a mandatory step in any time series prediction, since the obtained knowledge (regarding data distribution, stationarity, predictability, etc.) can be used in designing appropriate predictors. The prediction accuracy of trivial and linear predictors provides accuracy lower bounds for reasonable nonlinear prediction systems. Hence, any nonlinear predictor whose prediction accuracy does not exceed that of the previously mentioned predictors should be disregarded.

The results showed that: (1) the interpolation process altered the original time series data distribution; (2) both the original and the interpolated time series were non-stationary; (3) the price prediction was considerably easier than the price change prediction; (4) trivial and linear predictors provided better $nRMSE$ values than some nonlinear competition entries; (5) the directional symmetry measure was un-informative due to the properties of the interpolated time series.

The performed analysis suggests the following testbed design recommendations for future financial forecasting competitions: (i) it should be tested whether the time series is non-stationary and if it is, then model retraining should be allowed; (ii) a uniformly sampled financial time series should be selected as a testbed, or at least it should be a time series in which the interpolation process preserves the real-life data distribution; (iii) an exploratory data analysis should investigate the prediction accuracy as a function of prediction horizon in order to formulate a challenging but feasible forecasting problem; (iv) explicit rules (e.g. $nRMSE$ computed either on price or price change) for evaluating a predictor's accuracy should be provided, since predictors can be optimized differently for specific objectives.

References

[1] Box, G. and Jenkins, G. [1976] *Time Series Analysis. Forecasting and Control,* Prentice Hall.

[2] Caldwell, R. B. [1995] "Performance Metrics for Neural Network-based Trading System Development," *NeuroVe$t Journal,* Vol. 3, No. 2, pp. 13-23.

[3] Drossu, R. and Obradovic, Z. [1996] "Rapid Design of Neural Networks for Time Series Prediction," *IEEE Computational Science and Engineering,* Vol. 3, No. 2, pp. 78-89.

[4] Drossu, R. and Obradovic, Z. [1997] "An Analysis of the INFFC Cotton Futures Time Series: Lower Bounds and Testbed Design Recommendations," in *Nonlinear Financial Forecasting: Proceedings of the First INFFC,* Finance & Technology Publishing, pp. 241-261.

[5] Papoulis, A. [1984] *Probability, Random Variables, and Stochastic Processes,* McGraw-Hill.

[6] Press, W. H. et al. [1992] *Numerical Recipes in C,* Cambridge University Press.

[7] Tenorio, M. F. and Caldwell, R. B. [1996] "INFFC Update," *NeuroVe$t Journal,* Vol. 4, No. 1, pp. 7-9.