

An Analysis of the INFFC Cotton Futures Time Series: Lower Bounds and Testbed Design Recommendations

Radu Drossu & Zoran Obradovic
rdrossu@eecs.wsu.edu & zoran@eecs.wsu.edu

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

Abstract

This article reports 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 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, on the INFFC time series, determined in order to establish accuracy lower bounds for reasonable nonlinear prediction systems, identifies competition entries with prediction accuracies below the provided bounds. Outlier removal preprocessing did not show statistically significant accuracy improvements for trivial and linear predictors on the competition series. Finally, testbed design recommendations for future financial time series competitions are extracted from the results of this analysis.

1 Introduction

Financial time series are the result of complex and insufficiently understood interdependencies in fairly efficient markets. Consequently, financial markets forecasting models consider incomplete information, while factors not included in the models act as noise. In financial markets, it is not very likely to identify linear relationships, since these are relatively easy to discover and hence disappear fast when exploited. However, potentially profitable nonlinear data dependencies might exist even in the financial markets with large trading volume, since such relationships are difficult to discover due to the inherent complexity of nonlinear optimization techniques (huge computational requirements, presence of multiple local minima in the cost function etc.). Nevertheless, nonlinear optimization techniques have recently been used for

financial time series forecasting ranging from option pricing (Hutchinson et al. [1994]), corporate bond rating (Moody and Utans [1994]), stock index trading (Chenoweth and Obradovic [1996]) to currency exchange (Abu-Mostafa [1995]).

A serious drawback in comparing different financial markets forecasting techniques is a lack of standard benchmark problems. Typically, forecasting results are reported in the context of a particular data segment from a single time series, with a specific prediction objective, thus making the comparison of different forecasting techniques difficult.

The Santa Fe time series forecasting competition (Weigend and Gershenfeld [1993]) was one of the first successful attempts of evaluating time series forecasting models on pre-specified benchmark problems from a variety of domains, including also the financial domain of currency exchange rate forecasting. The First International Nonlinear Financial Forecasting Competition (INFFC) is a continuation of this standardization effort, that concentrates on financial markets forecasting only.

The INFFC benchmark data (Tenorio and Caldwell [1996]) was a cotton futures intra-day time series comprising 107,386 6-tuples, each providing 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. The competing systems had to provide forecasts for two different prediction horizons, 120 minutes and 1 day ahead, respectively. The real-life data was sampled non-uniformly, resulting in missing information for some minutes, but the INFFC rules clearly specified that an interpolation has to be performed to extend the data to all minutes of every trading day in order to facilitate the testing process. An interpolation algorithm for obtaining a uniformly sampled data set was provided to the competitors, according to which all missing price values are obtained by repeating the last available closing price. The interpolation process extended the 80,000 samples data set to approximately 261,000 used by the competitors and the 27,386 samples data set to approximately 67,000 samples used by the evaluation panel.

The objective of this article is to analyze the competition cotton time series and to evaluate the performance of simple predictors in order to establish accuracy lower bounds for reasonable nonlinear prediction systems. An additional goal is to provide testbed design recommendations for future financial forecasting competitions. 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.

Section 2 contains a statistical data analysis for the INFFC cotton futures time series. Section 3 reports the performance of trivial and linear predictors on the competition problem, whereas Section 4 explores the effect of time series outlier removal on prediction quality. Finally, Section 5 contains conclusions, as well as testbed design recommendations for future competitions based on the INFFC experience.

2 Statistical Analysis of the Competition Data

When forecasting the competition time series, different prediction goals can be imposed. One can attempt to predict an actual closing price $a(t+h)$ or the change in closing price $a(t+h)-a(t+h-1)$, where t is the current time and h is the desired prediction horizon (120 minutes or 1 day ahead in this competition). Although predicting the change in closing price is typically a considerably more difficult problem, it represents a more common financial forecasting goal. Unfortunately, the INFFC call for participation did not explicitly specify that the goal was to predict the change in closing price as opposed to the actual closing price. For that matter, this article considers both the actual price as well as the price change time series forecasting. The daily closing price, as well as the daily closing price changes with respect to the previous minute for the entire competition data are shown in Figs. 1 and 2, respectively. A zoom-in on the price and price change time series for a five day period and one minute resolution is shown in Figs. 3 and 4.

An analysis is performed in order to determine:

- whether the interpolation process explained in the Introduction alters the data distribution;
- 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*);
- 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 (Press et al. [1992]).

In the *chi-square test*, the data range of the two data sets to be compared is divided into a number of intervals (bins). Assuming that R_i and S_i represent the number of data samples in bin i for the first and the second data set, respectively, the chi-square statistic computes

$$\chi^2 = \sum_i \frac{(R_i - S_i)^2}{R_i + S_i},$$

with the sum taken over all bins. The complement of the incomplete gamma function,

$$Q(\nu, \chi^2) = \frac{1}{\Gamma(\nu)} \int_{\chi^2}^{\infty} e^{-t} t^{\nu-1} dt,$$

where

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt,$$

is then evaluated and a small value of Q (close to 0) indicates that it is unlikely that the two distributions are the same. Here, ν represents the number of degrees of freedom, which in the case when the two sets have the same number of data samples $\left(\sum R_i = \sum S_i\right)$, equals the number of bins minus one. If the previous restriction is not imposed, then ν equals the number of bins.

The Kolmogorov-Smirnov (K-S) test measures the absolute difference between two cumulative distribution functions S_{N_1} and S_{N_2} with N_1 and N_2 data points, respectively. The K-S statistic computes

$$D = \max_{-\infty < x < \infty} |S_{N_1}(x) - S_{N_2}(x)|.$$

The function Q_{KS} , defined as

$$Q_{KS}(\lambda) = 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2\lambda^2},$$

is computed for

$$\lambda = D \left(\sqrt{N_e} + 0.12 + 0.11 / \sqrt{N_e} \right),$$

where N_e is the effective number of data points computed as

$$N_e = \frac{N_1 N_2}{N_1 + N_2}.$$

A small value of Q_{KS} (close to 0) indicates that it is unlikely that the two distributions are the same.



Fig. 1: Daily Closing Price

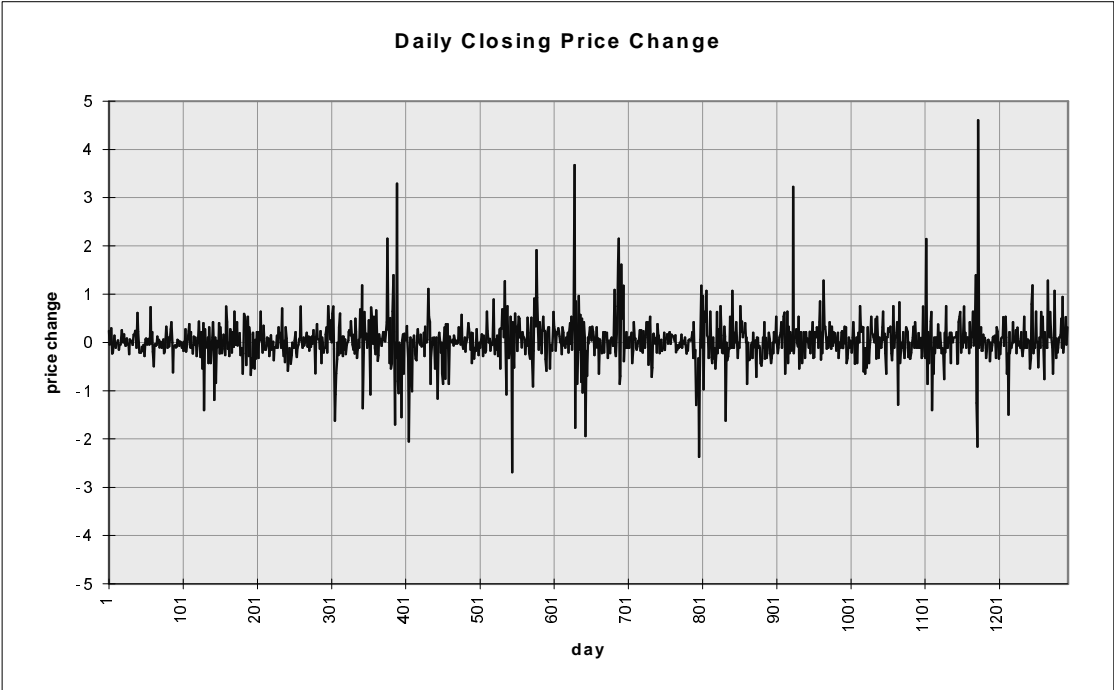


Fig. 2: Daily Closing Price Change with Respect to the Previous Minute



Fig. 3: Minute Closing Price for First Five Days

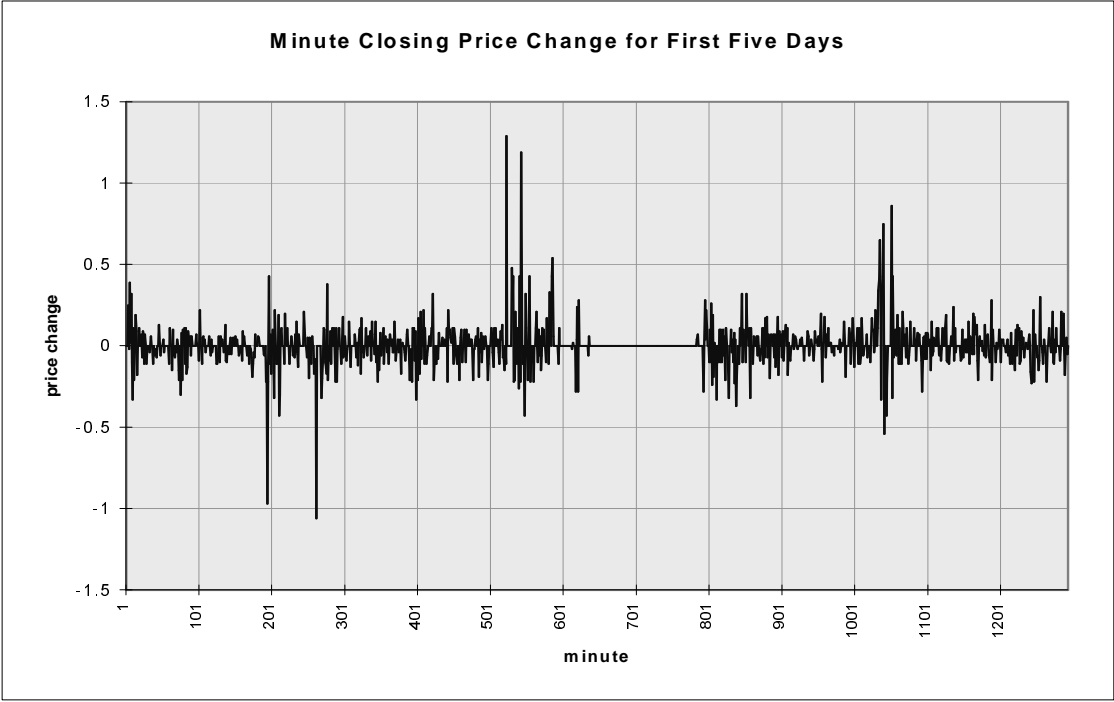


Fig. 4: Minute Closing Price Change for First Five Days

2.1 Interpolation Effects on Data Distribution

Normalized histograms for un-interpolated and interpolated training set prices are shown in Figs. 5 and 6, respectively, whereas the normalized histograms for un-interpolated and interpolated test set prices are shown in Figs. 7 and 8, respectively. The visual inspection of Figs. 5 and 6 indicates radically different distributions, while this is not obvious for Figs. 7 and 8.

More objectively, both the chi-square and the Kolmogorov-Smirnov tests yielded a probability 0.999999 of rejecting the null hypothesis that the un-interpolated and the interpolated price training data distributions are the same. Similarly, the chi-square and the Kolmogorov-Smirnov tests yielded a 0.999999 probability of rejecting the null hypothesis that the un-interpolated and the interpolated price test data distributions are the same, hence explicitly showing that the interpolation process altered the data distribution.

A similar analysis on price changes shows histograms of both interpolated training and interpolated testing data sets (Figs. 10 and 12) as being significantly more leptokurtic (pointed) than the corresponding histograms of the un-interpolated data sets (Figs. 9 and 11). This is due to the fact that a fairly large amount of data is introduced through the interpolation process by repeating the last available closing price when data is missing. This results in a large number of zero-valued price changes shown in the interpolated data histograms as a long bar centered about the zero value. This visual finding that the interpolation alters the price change distribution is confirmed by chi-square and Kolmogorov-Smirnov tests that both reject the null hypothesis with probabilities between 0.93 and 0.999999.

2.2 Stationarity Analysis

A *non-stationary* time series can be described as a time series whose characteristic parameters change over time. Common concepts include strict-sense, wide-sense, n-th order and weak-sense stationary processes (Papoulis [1984]). In general, non-stationarity detection can be reduced to identifying two sufficiently long, distinct data segments that have significantly different statistics (distributions). For non-stationary domains, the *single model* technique of building a prediction model on a certain data segment and using it for all subsequent predictions is usually inadequate, while better results can be obtained if retraining the model when significant changes in the distribution are signaled (Drossu and Obradovic [1996a]).

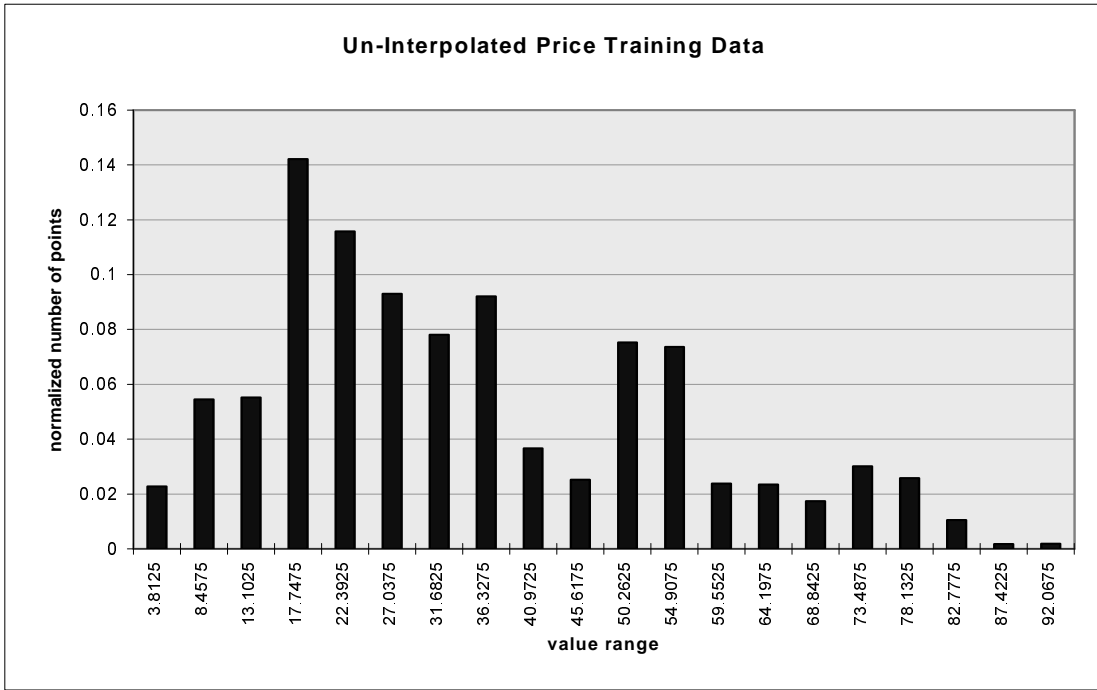


Fig. 5: Histogram for Un-interpolated Price Training Set

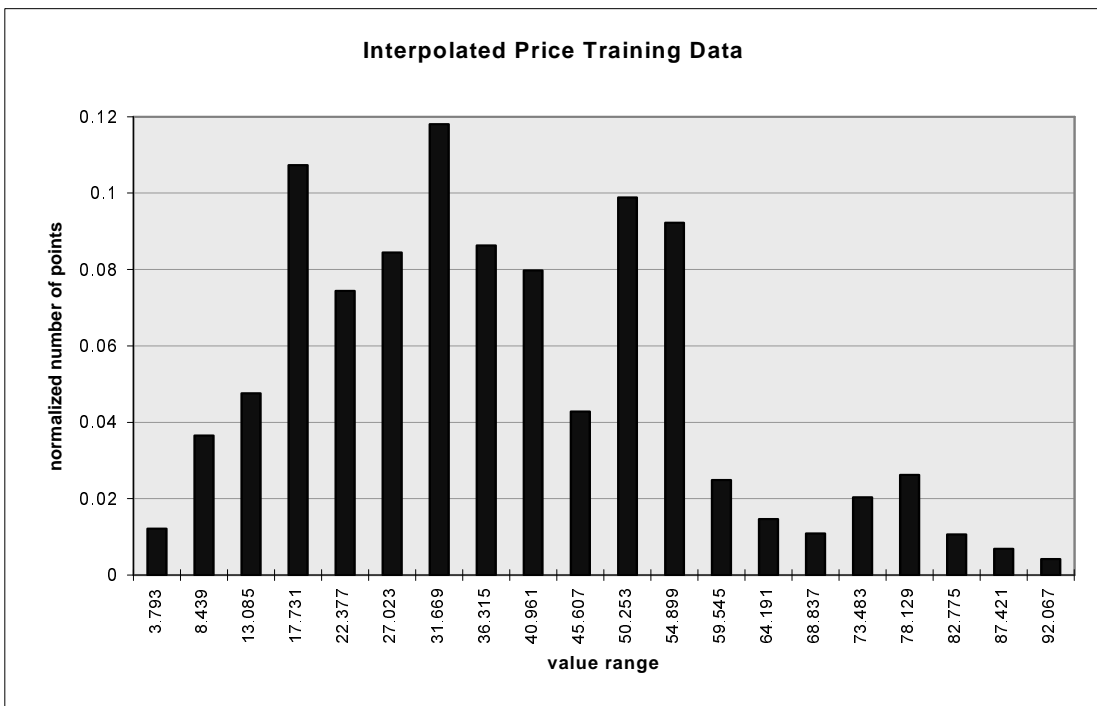


Fig. 6: Histogram for Interpolated Price Training Set

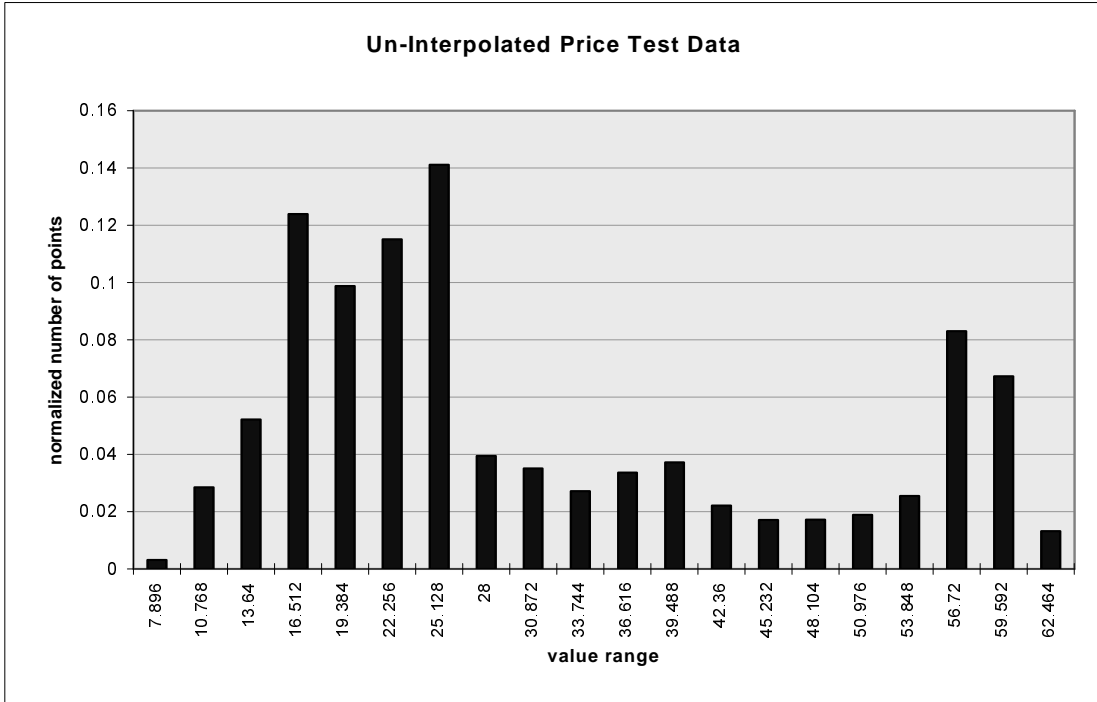


Fig. 7: Histogram for Un-interpolated Price Test Set

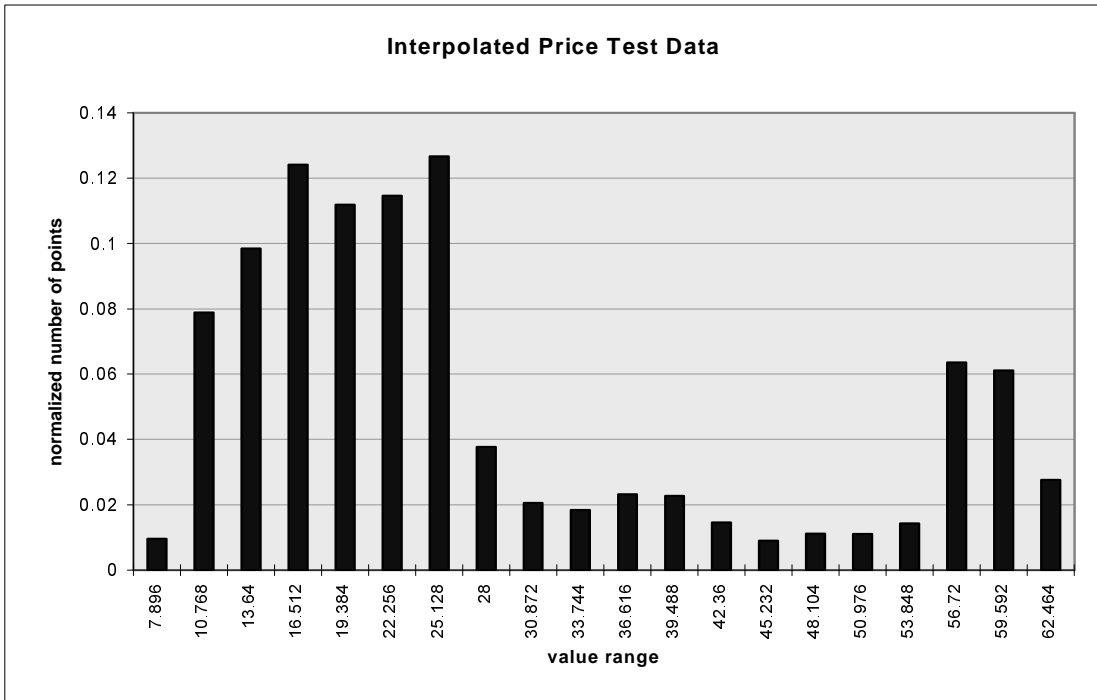


Fig. 8: Histogram for Interpolated Price Test Set

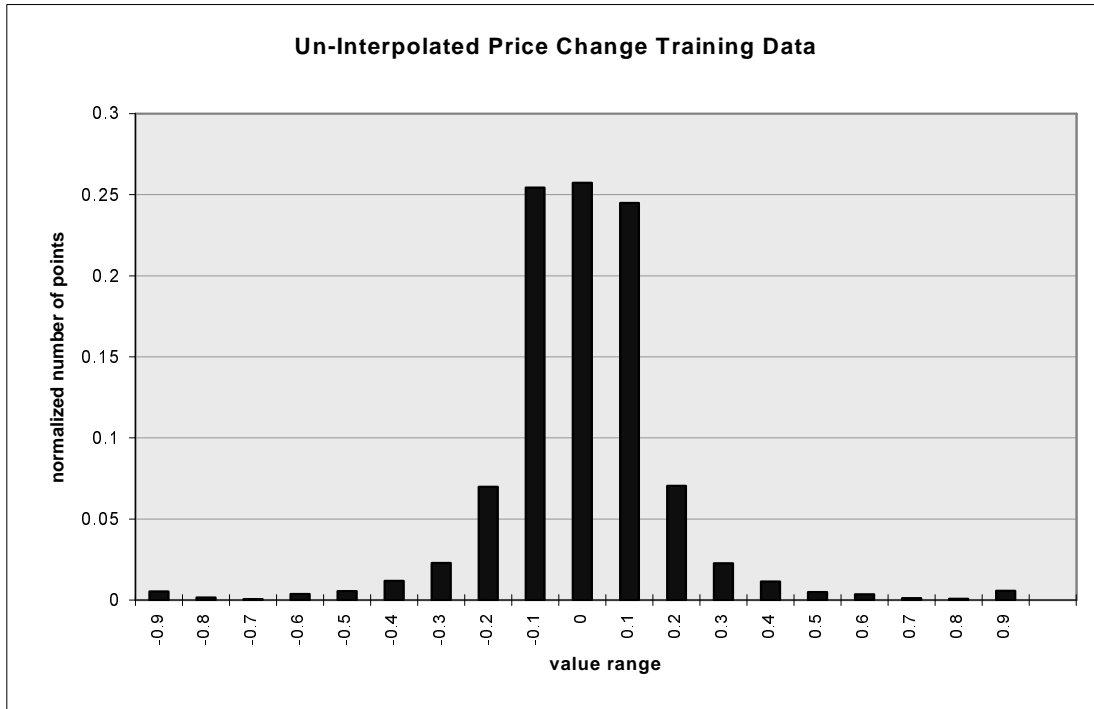


Fig. 9: Histogram for Un-interpolated Price Change Training Set

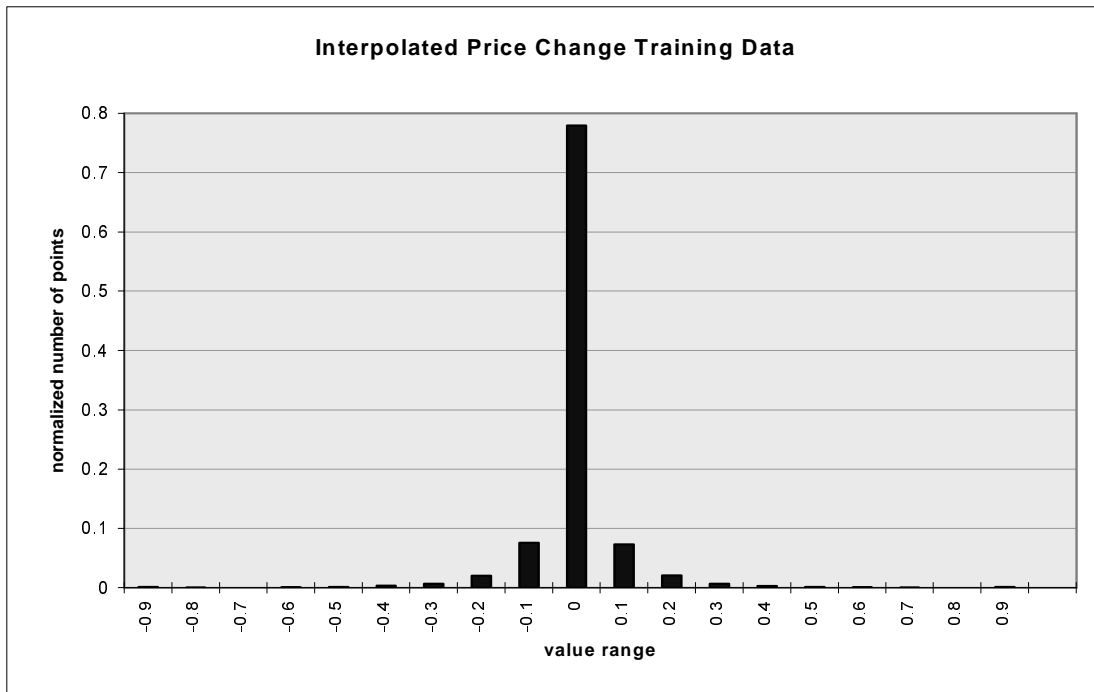


Fig. 10: Histogram for Interpolated Price Change Training Set

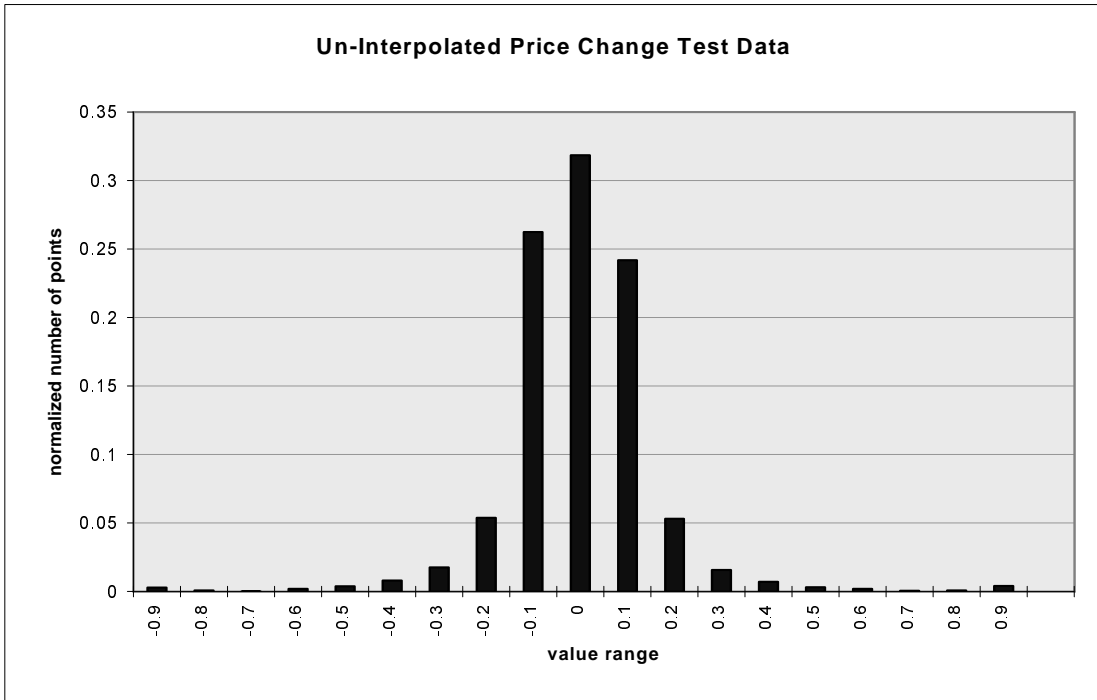


Fig. 11: Histogram for Un-interpolated Price Change Test Set

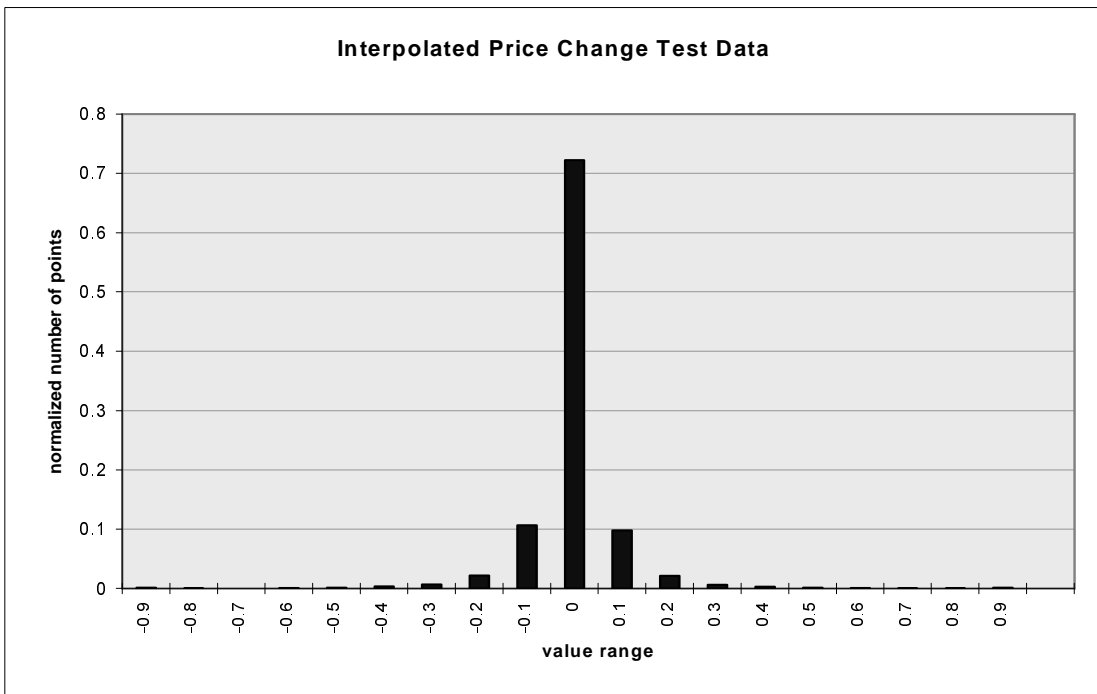


Fig. 12: Histogram for Interpolated Price Change Test Set

For the competition data, stationarity analysis is reduced to comparing the interpolated training and test distributions for the price, as well as for the price change time series. The histograms shown in Figs. 6 and 8 indicate significantly different price distributions, confirmed also by the chi-square and Kolmogorov-Smirnov tests that reject the null hypothesis, according to which the two distributions are the same, with probability 0.999999. Although the histograms shown in Figs. 10 and 12 for the interpolated price change training and test data sets are fairly similar, the chi-square and Kolmogorov-Smirnov tests reject the hypothesis that the distributions are the same with probabilities 0.999682 and 0.999999, respectively.

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.

2.3 Data Correlation

Autocorrelation plots for a 300 samples data segment from the training and the test parts of the interpolated price time series are shown in Figs. 13 and 14. The relatively slow drop of the autocorrelation function suggests a potentially large reliable prediction horizon when predicting actual prices.

Similar autocorrelation plots for the interpolated price change time series (shown in Figs. 15 and 16) suggest very short reliable horizon for price change prediction when using previous price changes only.

3 Trivial and Linear Predictors

Assuming the availability of reasonably 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.



Fig. 13: Autocorrelation for Interpolated Price Training Set

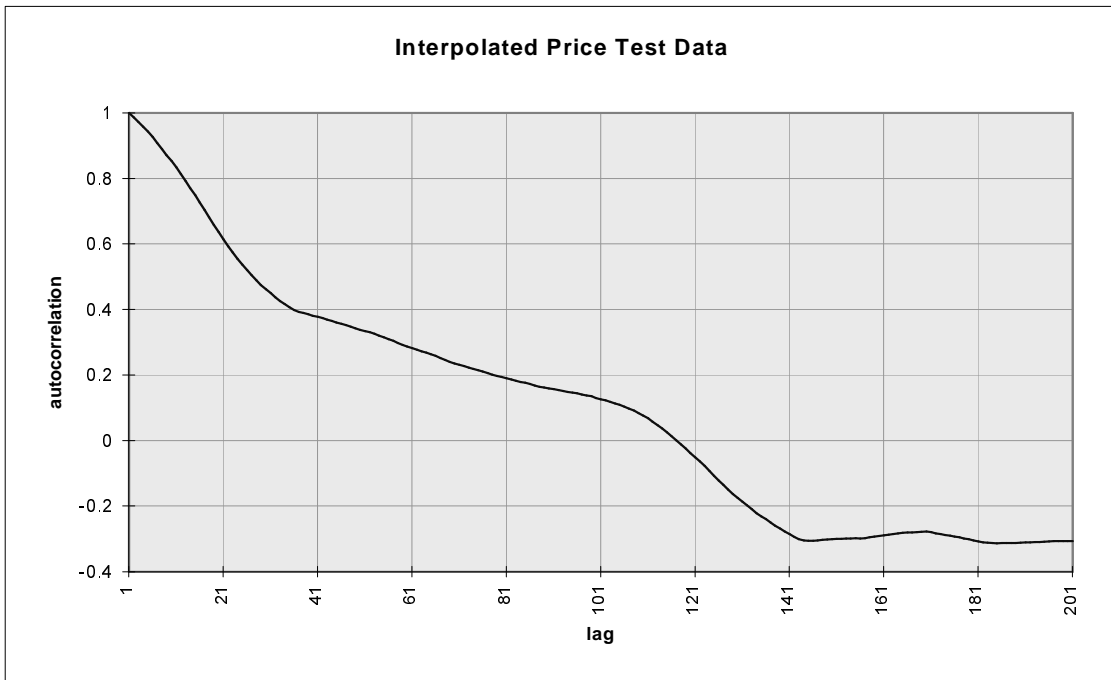


Fig. 14: Autocorrelation for Interpolated Price Test Set

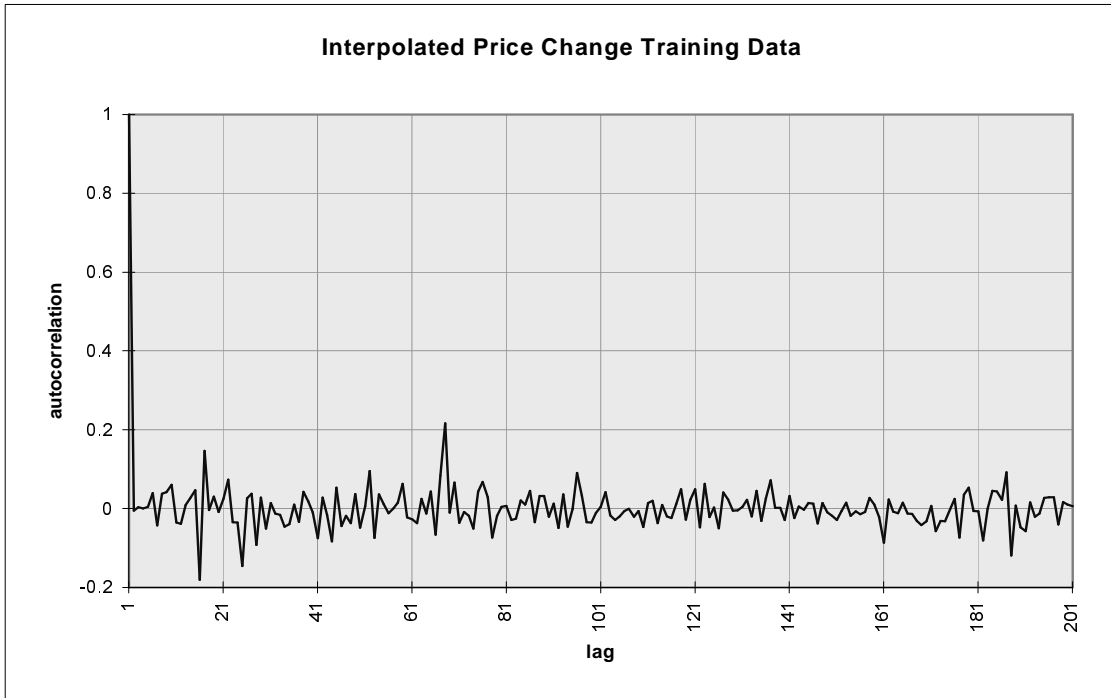


Fig. 15: Autocorrelation for Interpolated Price Change Training Set

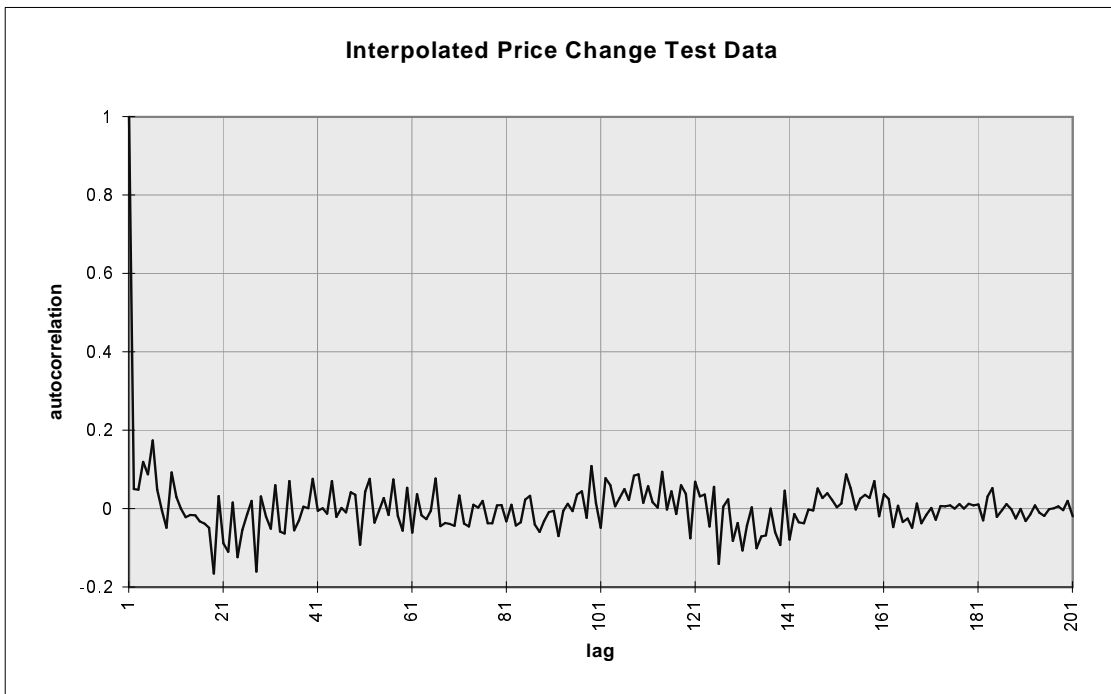


Fig. 16 Autocorrelation for Interpolated Price Change Test Set

A general linear time series model (Box and Jenkins [1976]) is the autoregressive moving average of orders p and q (ARMA(p,q)). It describes the process value as a weighted sum of p previous process values and the current, as well as q previous values of a random process. Formally, for a zero mean process, the ARMA(p,q) model for $\{x_t\}$ is given as

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots + \psi_q a_{t-q},$$

where $x_{t-1}, x_{t-2}, \dots, x_{t-p}$ represent the process values at p previous time steps, $a_t, a_{t-1}, \dots, a_{t-q}$ are the current and the q previous values of a random process, usually emanating from a normal (Gaussian) distribution with zero mean and $\phi_1, \dots, \phi_p, \psi_1, \dots, \psi_q$ are the model parameters.

The ARMA(p,q)-based predictor approximates the real process value x_t by a predicted value \hat{x}_t computed as

$$\hat{x}_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots + \psi_q a_{t-q}.$$

The error between the real process value x_t and the predicted value \hat{x}_t is the *residual* a_t .

The AR(p) model considered in this analysis is a special case of the ARMA(p,q) model described as

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + a_t.$$

The analysis also considers the autoregressive-integrated ARI(p) model which is an AR(p) model applied to differenced data.

A multitude of accuracy measures can be considered in order to evaluate the accuracy of a given predictor (Caldwell [1995]). However, many of these measures are either redundant (e.g. normalized root mean squared error and coefficient of determination are highly related) or encompassed in more powerful measures (e.g. normalized root mean squared error is more relevant than the root mean squared error). The only two standard 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 the directional symmetry (*DS*) defined as

$$nRMSE = \frac{\sqrt{\frac{1}{n} \sum_{t=1}^n (x_t - \hat{x}_t)^2}}{\sqrt{\frac{1}{n-1} \sum_{t=1}^n (x_t - \bar{x}_t)^2}},$$

where

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t,$$

and

$$DS = \frac{100}{n} \sum_{t=1}^n d_t,$$

where

$$d_t = \begin{cases} 1, & \text{if } (x_t - x_{t-1})(\hat{x}_t - \hat{x}_{t-1}) > 0 \\ 0, & \text{otherwise.} \end{cases}$$

The $nRMSE$ measure is always non-negative with smaller values indicating a better predictor. The DS measures the percentage of correctly predicted market directions, with larger values suggesting a better predictor, while $DS=50\%$ meaning that the market direction is predicted correctly for half of all predictions. Unfortunately, the DS defined as previously is meaningless for a time series with a large number of equal consecutive value pairs, since the DS accounts only for correctly predicted upward and downward trends. This problem is particularly serious for the interpolated competition time series, in which there are many missing values which are filled-in by replicas of the last available actual data. Consequently, we propose a modified directional symmetry ($modDS$) which takes into consideration all the correctly predicted directions (upward, downward, and no change), as well as computer truncation errors, defined as

$$DS = \frac{100}{n} \sum_{t=1}^n c_t,$$

where

$$c_t = \begin{cases} 1, & \text{if } (x_t - x_{t-1})(\hat{x}_t - \hat{x}_{t-1}) > 0 \text{ and } |x_t - x_{t-1}| > \varepsilon \text{ and } |\hat{x}_t - \hat{x}_{t-1}| > \varepsilon \\ & \text{or} \\ & |x_t - x_{t-1}| < \varepsilon \text{ and } |\hat{x}_t - \hat{x}_{t-1}| < \varepsilon \\ 0, & \text{otherwise,} \end{cases}$$

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

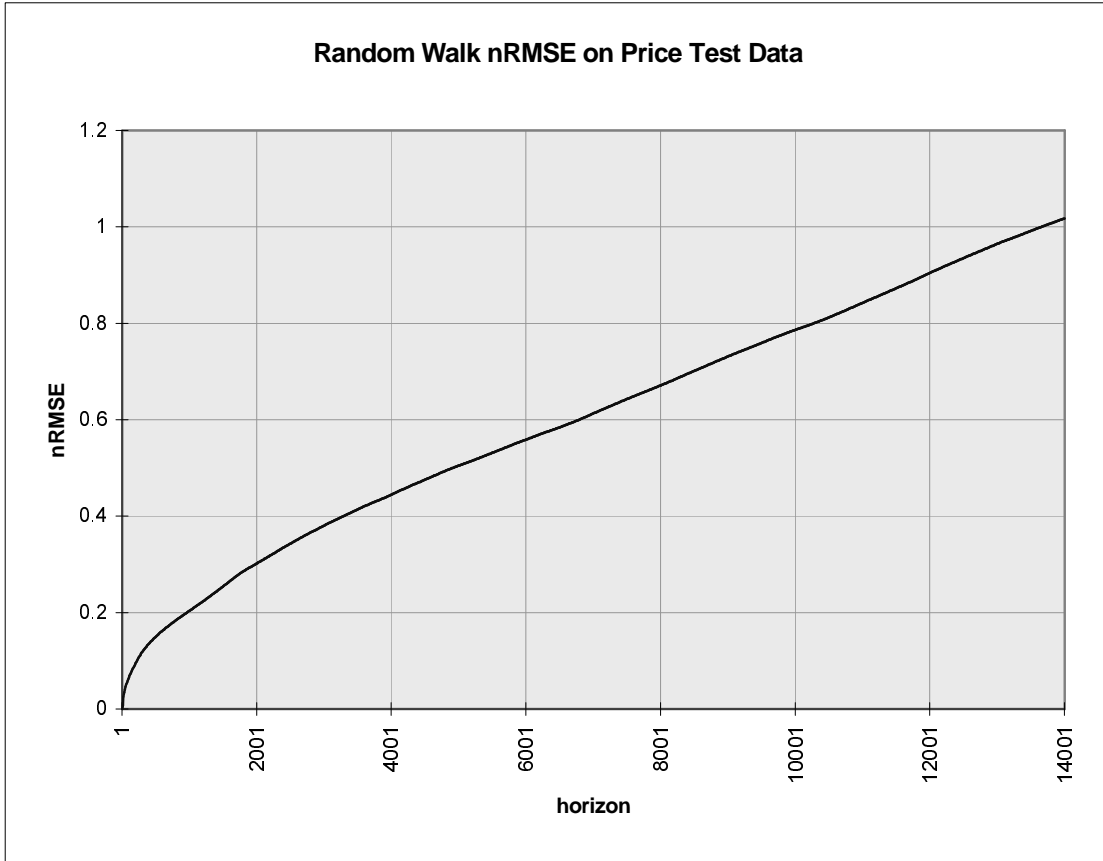


Fig. 17: Random Walk nRMSE vs. Prediction Horizon for Price Test Data

The $nRMSE$ as a function of prediction horizon for a trivial random walk predictor on the price test series is shown in Fig. 17. The relatively slow rise of the curve suggests a fairly easy prediction problem both for 120 minutes and 1 day prediction horizons, confirmed by the $nRMSE$ values shown in Tables 1 and 2. For the price prediction one can observe that the random walk predictor yields the best $nRMSE$ values both compared to the mean predictor and to the linear autoregressive models. It should be noted that the search for appropriate $AR(p)$ and $ARI(p)$ models was restricted to orders $p \leq 10$ and is using a sampling rate of one minute. Consequently, these linear models do 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 $nRMSE$ values for the mean predictor are extremely poor, its $modDS$ values for price prediction are 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.

| Predictor | <i>DS</i> | <i>modDS</i> | <i>nRMSE</i> |
|-------------|-----------|--------------|--------------|
| 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 Prediction 120 Minutes ahead

| Predictor | <i>DS</i> | <i>modDS</i> | <i>nRMSE</i> |
|-------------|-----------|--------------|--------------|
| Random Walk | 8.789 | 54.955 | 0.113 |
| Mean | 0.000 | 64.255 | 1.161 |
| AR(3) | 14.014 | 43.416 | 0.387 |
| ARI(1) | 8.897 | 53.633 | 0.394 |

Table 2: Price Prediction 1 Day ahead

| Predictor | <i>DS</i> | <i>modDS</i> | <i>nRMSE</i> |
|-------------|-----------|--------------|--------------|
| Random Walk | 15.377 | 44.418 | 1.413 |
| Mean | 0.000 | 49.397 | 1.000 |
| AR(3) | 20.646 | 40.995 | 1.338 |
| ARI(1) | 16.198 | 43.458 | 1.399 |

Table 3: Price Change Prediction 120 Minutes ahead

| Predictor | <i>DS</i> | <i>modDS</i> | <i>nRMSE</i> |
|-------------|-----------|--------------|--------------|
| 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 4: Price Change Prediction 1 Day ahead

The price change prediction results are presented in Tables 3 and 4. As expected, the *nRMSE* values for the random walk predictor are significantly larger as compared to those obtained on the price prediction, confirming the increased difficulty of the problem. However, the *nRMSE* values for the mean predictor improve for the price change prediction since the mean of the price change training series is the same as the mean of the price change test series, while this is 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 is 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 appears to be better than the random walk and the AR predictors with respect to both *nRMSE* and *modDS*.

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 (Drossu and Obradovic [1996b]). However, this was not necessary for this analysis, since even these simple predictors were able to outperform some of the competition entries (Tenorio and Caldwell [1996]).

4 Outlier Removal Effects on Trivial and Linear Predictors

The existence of outliers makes time series prediction particularly challenging. In practice, the outliers are difficult to predict due to their relative sparsity in the training set, while their existence in the training set negatively affects the optimization process for the remaining data. An improved prediction accuracy using trivial or linear predictors on data with outliers removed would provide strong evidence in favor of outlier removal as a preprocessing step for nonlinear predictors. However, outlier removal should still be considered for nonlinear predictors even if its advantages are not evident for trivial and linear predictors.

The outliers can either be removed from the data set or replaced by appropriately filtered values. A common outlier removal technique, called in this article STDO, consists of removing all the process values that are more than s standard deviations away from the data mean, with typical values for the dispersion threshold s being 2 and 3. In the experiments reported in this section, the identified outliers were replaced by the last observed (non-outlier) process value.

A block diagram of an alternative outlier removal technique proposed in this article, called MFO, based on median filtering is shown in Fig. 18. The idea of this technique is to either leave the original process value unchanged, if the value is not an outlier, or to replace it by a value obtained through median filtering, if it is an outlier. The median filtering assumes that a window containing $2k+1$ samples, $x_{t-k}, \dots, x_{t-1}, x_t, x_{t+1}, \dots, x_{t+k}$, slides over the data set replacing the x_t value by m_t which is the $(k+1)$ -st largest value in the current window. The MFO technique computes the difference δ_t between the actual process value x_t and the corresponding median filtered value m_t and also the mean μ and the standard deviation σ of the $\{\delta_t\}$ series. On a given x_t , the output of MFO y_t is computed as

$$y_t = \begin{cases} x_t, & \text{if } \mu - s\sigma \leq \delta_t \leq \mu + s\sigma \\ m_t, & \text{otherwise,} \end{cases}$$

where s is a prespecified dispersion threshold.

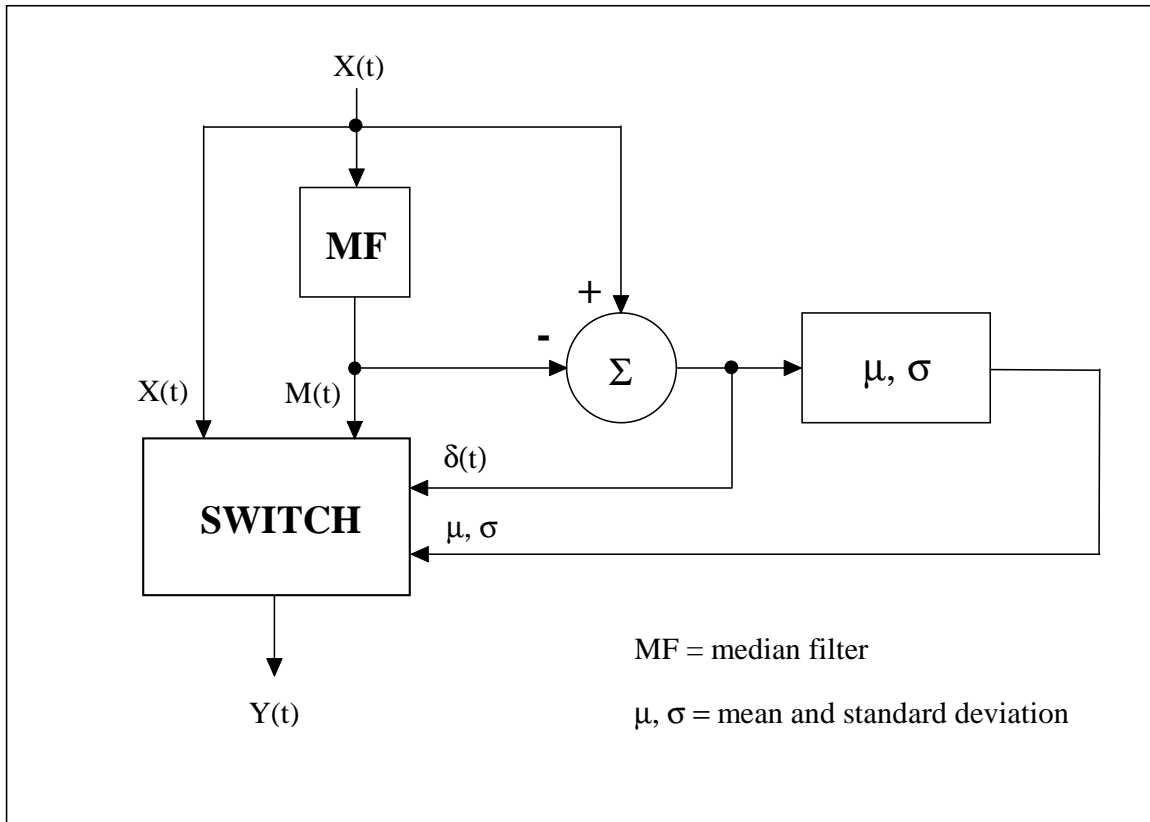


Fig. 18: MFO Outlier Removal Technique

The STDO and MFO outlier removal experiments were performed using a dispersion threshold $s=2$, whereas the window size for the median filter was set to five ($2k+1=5$). STDO removed 5.1% training set values and 4.0% test set values, whereas MFO removed 1.8% training set values and 3.4% test set values. However, the outlier removal did not lead to significantly better prediction accuracy neither for the trivial (random walk and mean predictors) nor for the autoregressive predictors. Obviously, outlier removal with $s>2$ would eliminate a subset of the data removed using $s=2$, resulting in similar prediction accuracy as for $s=2$. On the other hand, the STDO and the MFO techniques with $s=2$ changed the data distribution considerably, as confirmed by the chi-square and Kolmogorov-Smirnov tests, thus indicating that outlier removal using $s<2$ would result in a significant information loss.

5 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. In addition, the effects of outlier removal on the predictability of the competition time series were investigated.

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 outlier removal is important, since outliers are difficult to predict due to their relative sparsity in the training set, while their existence in the training set negatively affects the optimization process for the remaining data.

The results show that:

- The interpolation process altered the original time series data distribution.
- Both the original and the interpolated time series were non-stationary.
- The price prediction was considerably easier than the price change prediction.
- Trivial and linear predictors provided better $nRMSE$ values than some nonlinear competition entries.
- 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:

- It should be tested whether the time series is non-stationary and if it is, then model retraining should be allowed.
- 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.
- 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.
- 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] Abu-Mostafa, Y. [1995] "Hints," *Neural Computation*, Vol. 7, No. 4, pp. 639-671.

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

- [3] Caldwell, R. B. [1995] "Performance Metrics for Neural Network-based Trading System Development," *NeuroVe\$t Journal*, Vol. 3, No. 2, pp. 13-23.
- [4] Chenoweth, T. and Obradovic, Z. [1996] A Multi-Component Nonlinear Prediction System for the S&P 500 Index," *Neurocomputing*, Vol. 10, No. 3, pp. 275-290.
- [5] Drossu, R. and Obradovic, Z. [1996a] "Regime Signaling Techniques for Non-stationary Time Series Forecasting," *NeuroVe\$t Journal*, Vol. 4, No. 5, pp. 7-15.
- [6] Drossu, R. and Obradovic, Z. [1996b] "Rapid Design of Neural Networks for Time Series Prediction," *IEEE Computational Science and Engineering*, Vol. 3, No. 2, pp. 78-89.
- [7] Hutchinson, J. M. et al. [1994] "A Nonparametric Approach to Pricing and Hedging Derivative Securities Via Learning Networks," *The Journal of Finance*, Vol. 49, No. 3, 851-889.
- [8] Moody, J. and Utans, J. [1994] "Architecture Selection Strategies for Neural Networks: Application to Corporate Bond Rating Prediction," in *Neural Networks in the Capital Markets* (A. N. Refenes, editor), John Wiley & Sons.
- [9] Papoulis, A. [1984] *Probability, Random Variables, and Stochastic Processes. Second Edition*, McGraw-Hill.
- [10] Press, W. H. et al. [1992] *Numerical Recipes in C. Second Edition*, Cambridge University Press.
- [11] Tenorio, M. F. and Caldwell, R. B. [1996] "INFFC Update," *NeuroVe\$t Journal*, Vol. 4, No. 1, pp. 7-9.
- [12] Weigend, A. S. and Gershenfeld, N. A. (eds.) [1993] *Time Series Prediction: Forecasting the Future and Predicting the Past*, Addison-Wesley.