The elusive quest for preserved quantities in financial time series: making a case for intraday trading strategies

William Krivan
Working Paper 01-16
Krivan Capital LLC
Santa Monica, CA 90404, USA
william@krivan.com

April 22, 2016

Abstract

In the context of the supervised learning problem for time series forecasting, we focus on financial time series and use the currency pair EURUSD to highlight issues that arise when daily data are utilized for one-day forecasts of currency exchange rate moves. In light of our results for forecast horizons of one day or more, we take a closer look at the EURUSD time series data to get a better understanding of typical intraday moves and their magnitude and how their potential can be harnessed for the development of consistently profitable trading strategies. By combining the results of our own numerical studies with published findings from the literature and illuminating them from a practical perspective, we motivate a simple intraday trading strategy for EURUSD that avoids some of the problems associated with longer-term forecasts.

1 Introduction

There is a sizable body of literature describing quantitative approaches to the trading of financial instruments such as stock index futures or currencies (forex). Many existing methods employ supervised learning techniques, e.g., Support Vector Machines (SVMs [1]) or Artificial Neural Networks (ANNs [2]), which utilize historical data to train a model that is subsequently used to predict the future behavior over forecast horizons of one day or more [3, 4]. Despite their sophistication and their inherent classification performance, the approaches described in many publications face two serious problems whose origin lies in the nature of the financial time series data and that we discuss in the remainder of this section.

1.1 Training window size and concept drift

Frequently, training windows of a size of one year or more are used to train a model that is then utilized for predictions in a more recent time period ranging from days to years [5, 6, 7, 8, 9, 10, 11, 12, 13].

While it may generally be true for static data that a larger data set for training leads to better predictive models, the *concept drift* [14, 15] for time-dependent problems can lead to changing relationships between the features and the target variable (as well as between the features themselves)

as time progresses; therefore, shorter training windows, which more accurately reflect the current relationship between the features and the target variable, can produce better results. (See Appendix A for a very simple concept drift toy problem and Appendix B for an illustrative example involving the directional prediction of daily moves of the currency pair EURUSD.) For financial time series, this issue has been addressed in the literature [16]; but, in our opinion, it does not seem to be adequately reflected in many more recent publications. As for example discussed in [17], the problem of the concept drift can be tackled by the use of adaptive techniques or by attempting to find an optimal choice for the training window size.

1.2 Forecast horizon

In many publications, the forecast horizon is on the order of one day or more [6, 7, 8, 9, 10, 12]. If only daily data are used, this choice makes it impossible to harness detailed information about intraday moves in order to increase the predictive performance.

One may speculate that it may be worthwhile for a practitioner to explore shorter forecast horizons in conjunction with carefully chosen entry and exit times that — as we will show below for EURUSD — might be advantageous for the design of profitable trading methodologies. In the following, we would like to outline two problems with forecast horizons of one day or more that can be avoided by performing intraday forecasts with data of finer granularity.

First, as we demonstrate in Appendix B, forecasts over horizons of one day or more based on daily data are prone to sensitivity to small differences between sets of data from different sources. While differences between data sets of higher frequency also lead to inconsistencies between individual forecasts, the overall outcome for one day's worth of data is the integral over many shorter-term forecasts, resulting in a consistent picture. Additionally, we can focus on intraday periods of significant moves and liquid markets instead of being limited to daily close data that may have been recorded at inconsistent times during thin market conditions.

Second, by avoiding the use of daily data, we can bypass the problem of ambiguity with regard to the omission of holidays; it is difficult if not impossible to decide which days should be omitted from historical data, and the handling of holidays is not consistent between different data sources. This is a problem because we would like consistently continuous data over the entire span of the training and testing time frames.[1] In contrast, the omission of holidays constitutes a very different problem for the one-minute data that we utilize in our production environment: formally, the effect of the inclusion or omission of a day's worth of daily data corresponds to the inclusion or omission of one minute in our intraday approach. The latter would cause a problem when markets are thin (for example on a late Friday afternoon on the East Coast of the US): if we collected data from several days during those times, some data points would effectively be missing from the data set due to lack of market activity; consequently, the results would likely be erratic. As we show in the following, however, it is possible to focus on times when the continuity of the data flow is likely to be consistent for a currency of interest. Moreover, it turns out that there are patterns of intraday behavior that are preserved over periods of years with dramatically varying market conditions.

[1] For example, it is not clear why it should matter for a prediction of the one-day change of the EURUSD close from April 11^{th} to 12^{th}, 2016 that Easter Monday, March 28^{th}, 2016 was a market holiday in Europe. And, since many global markets (including forex) were open on Easter Monday, it is not clear why the daily data point for that day should not be included (i) in training the model and (ii) in the feature calculations for the April $11^{\text{th}}/12^{\text{th}}$ prediction.

2 A case for intraday trading strategies

In the appendix, we present a further discussion of the concerns regarding the training window size outlined in Section 1.1 above. Specifically, the numerical experiments described in Appendix B demonstrate that the choice of an appropriate size for the training data window is a non-trivial task. We also show that small differences between data sets from different sources can have a dramatic effect on the results obtained by predictive methods; thus, robustness is a serious issue.
Instead of discussing possible incremental computational improvements aimed at enhancing the accuracy of one-day predictions, however, the remainder of this section will focus on the characterization of intraday behavior to get a better understanding of typical intraday moves and their magnitude and how they could be harnessed for the execution of profitable trades.

2.1 Analysis of one-minute data for EURUSD

We have compiled data for EURUSD from July 2012 until March 2016 and will use the data to identify patterns that consistently appear during the entire time period studied. In conjunction with published findings from the literature for earlier years, we consider the time period of almost four years to provide sufficient evidence to suggest that there are intraday patterns of behavior that are preserved over periods of years with dramatically varying market conditions.
We obtained one-minute bid open-high-low-close data for EURUSD from Dukascopy Bank SA [18], as well as from OANDA Corporation [19], leading to identical conclusions. We excluded weekends from the data and used the Perl DateTime Module [20] to convert time stamps from Greenwich Mean Time (GMT) to Central European Time (CET) or Central European Summer Time (CEST, during the period of daylight saving).
For the intraday movement, we characterize volatility[2] in light of the question when, regardless of direction, we can expect a move that will be large enough to generate profit within a desired time interval after accounting for transaction costs. As a metric for the magnitude of intraday movement, we compute the maximum change (increase or decrease) from a given time t_0 over a k-minute interval ($k = 1, 2, 3, \ldots$). We define the maximum increase $\Delta y_{\max}$ of the exchange rate y over k minutes as

$$\Delta y_{\max} \equiv \max\{y_t - y_0\} \text{ where } t_0 \leq t \leq t_0 + k \,\text{minutes}\,, \tag{1}$$

and, analogously, the maximum decrease $\Delta y_{\min}$ as

$$\Delta y_{\min} \equiv \min\{y_t - y_0\} \text{ where } t_0 \leq t \leq t_0 + k \,\text{minutes}\,. \tag{2}$$

In order to represent the volatility as a function of the hour h of the day, we then define the mean maximum increase $\overline{\Delta y_{\max}}(h, k)$ as in the following example for $h = 10$ and $k = 20$:

$$\overline{\Delta y_{\max}}(h = 10, k = 20) \equiv \text{mean}\{\Delta y_{\max} \mid t_0 = 10{:}01, 10{:}02, \ldots, 11{:}00 \text{ and } k = 20\}\,. \tag{3}$$

The reason for t_0 ranging from $10:01$ to $11:00$ and not from $10:00$ to $10:59$ is that we select the time of the day based on the time stamps of the one-minute open-high-low-close data. For

[2] We use the term volatility as a loose synonym for "magnitude of variability." This is different from the definition of implied volatility that is derived from an option pricing model and that is the basis for the CBOE Volatility Index® [21].

each of the minutes with a given hourly time stamp, we then compute the maximum increase by determining the maximum relative to the open of the next minute; taking the maximum relative to the close values yields virtually identical results. The mean maximum decrease $\overline{\Delta y_{\min}}(h,k)$ is defined analogously to Equation (3).

For data from July to December 2015, Figure 1 shows the mean maximum increase $\overline{\Delta y_{\max}}$ as a function of both the hour h of the day (CET/CEST) and k. In the contour plot, the colors are separated by isolines and range from blue for $0 \leq \overline{\Delta y_{\max}} < 0.0001$ to red for $0.0018 \leq \overline{\Delta y_{\max}} < 0.0019$. (The values at the end of the day after the close of the North American markets should be viewed with a grain of salt: that period of the day is irrelevant for our methodology, which focuses on the European session; therefore, we can ignore weekend gaps, early market closing days, and other phenomena that distort the data late in the day.) We can see a monotonic increase of $\overline{\Delta y_{\max}}(h,k)$ as a function of k at a rate that is decreasing with k, and there are two ridges that correspond to two distinct maxima as a function of the hour h of the day that we discuss below. Analogously, the mean maximum decrease $\overline{\Delta y_{\min}}(h,k)$ (not shown) is monotonically decreasing with k and the contour plot would feature two deep valleys. In Appendix C, we give a plausible argument for this behavior that is based on a simple random walk model.

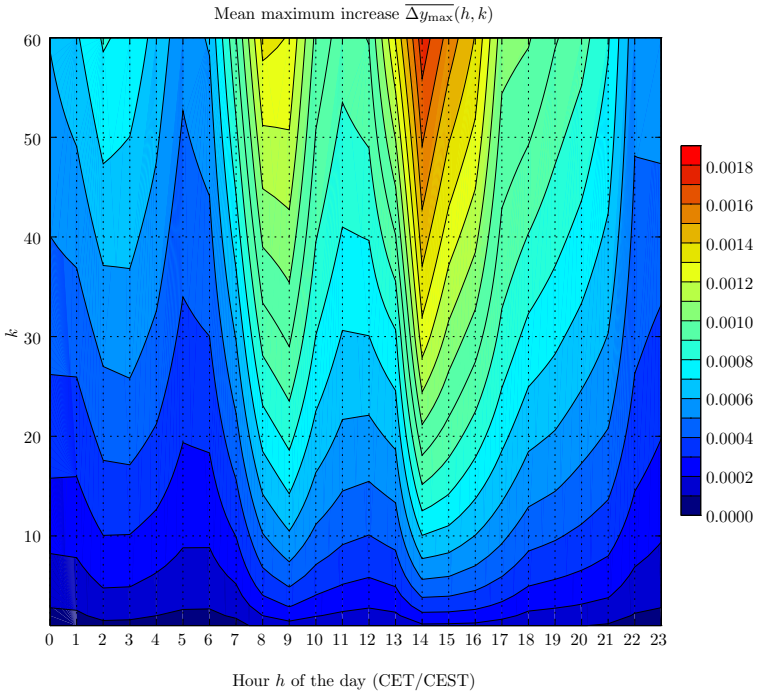


Figure 1: Mean maximum increase over k-minute intervals as a function of both the hour h of the day and k for EURUSD from July to December 2015. See text for details.

Figure 2 shows the mean (left) and median (right) maximum increase (top) and maximum decrease (bottom) for EURUSD for $k = 15$ minutes; choices of smaller or larger values of k yield a similar picture. The mean and median were computed from one-minute data for seven six-month time periods from July 2012 to December 2015 (H212, ..., H215), and one time period from January

2016 until March 25$^{\text{th}}$, 2016 (H116, incomplete at the time of this writing).
We see that our findings for the years 2012–2016 are largely consistent with the detailed analysis for 1999–2001 in [22]. (Cf. specifically Figure 3 therein, which shows the intraday activity of EURUSD for 1999–2001.) There are two distinct maxima: one that coincides with the European market open at around the hours 8–10 local time (CET/CEST) and another one during the overlap of the European and North American sessions at around the hours 14–16 CET/CEST. Furthermore, there is a less pronounced maximum marking the beginning of the Asian session that follows a trough after the North American session, and there are considerable drops during the lunch hours in Tokyo and, to a lesser extent, in Europe. (Within the scope of the present paper, we ignore inconsistencies due to different daylight saving schedules in North America and Europe.)

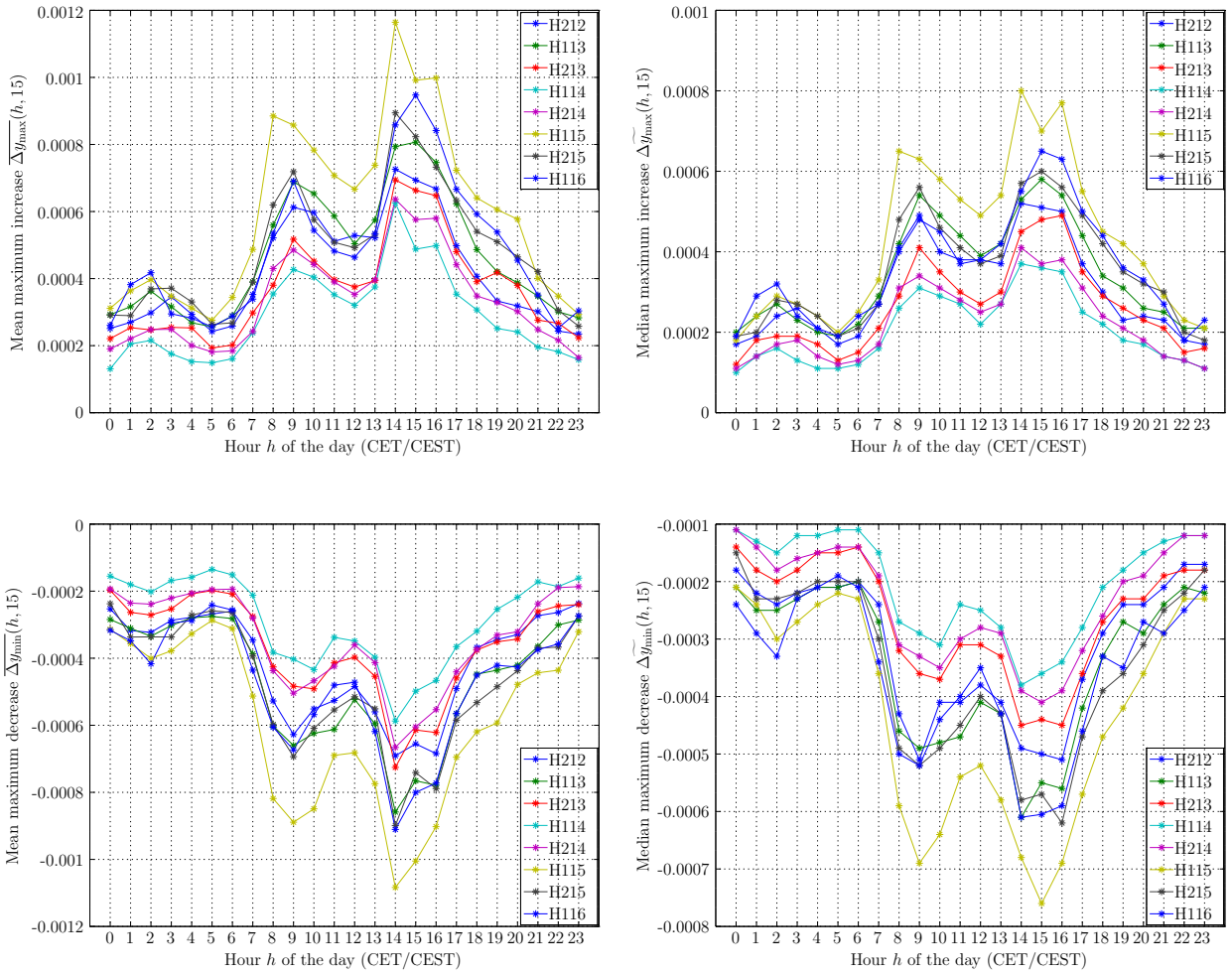


Figure 2: Extrema of changes of EURUSD over 15-minute intervals. See text for details.

It is remarkable that, qualitatively, this intraday pattern is preserved despite quantitative differences that reflect different market conditions during the different time periods. In Figure 2, we can for

example see low volatility in H114 compared to higher volatility in H115. This difference in volatility is qualitatively consistent with results that we obtained from an analysis of data of a different granularity that was based on very different methods: the analysis of daily data with Bollinger Bands, the Average True Range, and the Standard Deviation — standard technical indicators used as metrics for volatility [23] — is illustrated in Figure 3 and shows low volatility in H114 in comparison with H115.

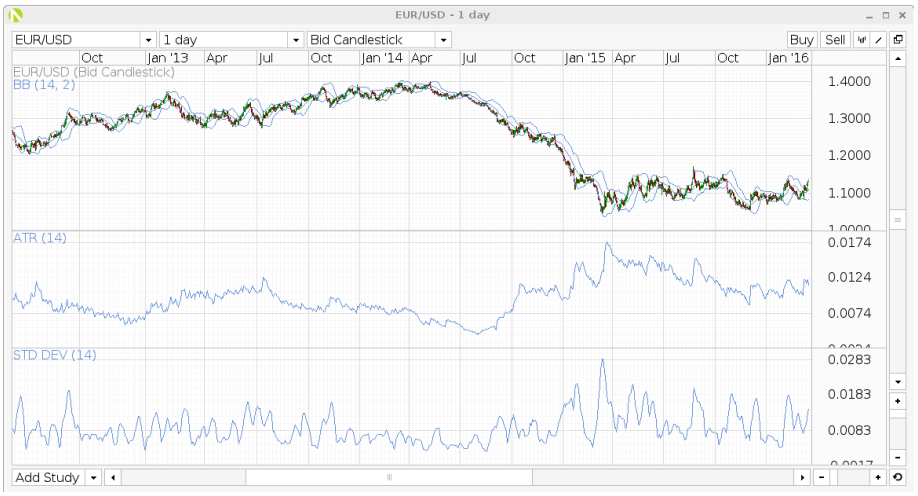


Figure 3: The daily chart for EURUSD from July 2012 until March 2016 is shown on the top panel along with the (14,2) Bollinger Bands. The 14-day Average True Range is shown in the center, and the 14-day Standard Deviation in the bottom part. See text for details. Data and graphing software have been obtained from OANDA Corporation [19].

2.2 Towards an intraday trading strategy

Based on the above findings, we can conjecture that promising trading activity of EURUSD should be concentrated during the European session because that is when significant moves can be expected. Assuming a bid-ask spread of approximately 0.0001, which is equivalent to the transaction costs per unit, Figures 1 and 2 suggest that holding periods of several minutes would be sufficient to achieve an average net profit if it were not for a slight caveat: the "only" thing we would need to figure out at any given time is if we should buy, sell, or sit still and do nothing. This is of course an utterly non-trivial task, and, instead of discussing supervised learning approaches and other techniques similar to the ones noted above for longer-term forecasts, we combine the results of our own numerical studies with published findings from the literature that lend themselves to a profitable — and surprisingly simple — trading strategy for EURUSD.

It has been shown that currencies tend to depreciate during local trading hours. To our knowledge, this time-of-day effect was first reported in [24], and subsequent studies [25] have shown in more detail that it appears to be stable over time. Data used in [25] range from 1997 to 2007 and show specifically for EURUSD that there is a significant depreciation of EURUSD during the European session that can be utilized for a trading strategy that is profitable even after accounting for transaction costs.

These findings form the basis for the approach used in our production environment. But the ideas

from the literature can be utilized even without further modifications: our numerical simulations (allowing for transactions every minute) have shown that, if considering short positions[3] only, a profitable trading strategy can be pursued for EURUSD by opening a EURUSD short position just before the start of the European session and closing the position at around the start of the North American session. For this purpose, we conducted a number of tests that included a Wilcoxon signed-rank test as an alternative to a paired t-test to compare the distribution of returns from short positions against the returns from positions that were opened and closed at the same times, but for which the trade direction had been chosen randomly.[4]
As we show in Appendix C, the specific times for optimal entry and exit for our short positions vary somewhat from [25]; but, our results, which we obtained for data from 2012 to 2016, are consistent with the results reported in [25] in the sense that, on average, we see a consistent decline of EURUSD during the European session.

3 Conclusions and outlook

By combining the results of our own studies with published findings from the literature, we lend support to a simple intraday trading strategy for EURUSD [25]. The strategy exploits the tendency of EURUSD to decline during the European session and avoids some of the problems and inconsistencies associated with longer-term forecasts that we discuss in the appendix. While a simple short-only strategy is at least marginally profitable on average for the time period considered in the present paper, there are sessions that yield an upward move of EURUSD and that therefore result in a loss. Our current work focuses on the identification of variables that can be used to characterize the "market state" before the start of the European session and thereby provide information about the expected directional move of EURUSD during European trading. In this manner, we have been able to further increase the accuracy of our predictions.

4 Supplementary files

The original daily data files that were downloaded from Dukascopy Bank SA and OANDA are available at `http://www.krivan.com/wp/wp0116` along with the further processed files. Furthermore, two simple scripts to generate the latter and a script to reproduce Figure 5 (b) are available as open-source software tools under the GNU General Public License. Please refer to the hosted documentation for further details.

Acknowledgments

We would like to thank Dukascopy Bank SA and OANDA Corporation for granting us permission to use their data for the purpose of this article. We would like to thank Indrė Žliobaitė for helpful suggestions and critical comments.

[3] A short position is opened when one expects an asset's value to decline; in this case it means selling EUR and buying the equivalent amount of USD, motivated by the expectation that the EURUSD exchange rate will decline.

[4] This is analogous to a drug study when the effect of the drug is assessed by comparing certain data from the same individuals before and after their treatment. We also conducted different numerical experiments with independent samples and performed a two-sample Kolmogorov-Smirnov test as well as a Wilcoxon rank-sum test, yielding identical conclusions for the optimal trade direction and time.

Appendix A Grapefruit, oranges, and lemons: a concept drift toy problem

Consider a citrus processing plant that receives from a specific supplier a mix of grapefruit, oranges, and lemons that needs to be separated for further processing in three different processing lines.
The sizes of different citrus fruit vary considerably depending on variety and provenance [26]; but, the task of separating the fruit into three classes could simply be performed by measuring their diameter if we, based on our knowledge from previous deliveries from the specific supplier, could assume that we know that the typical diameter of the grapefruit from the supplier is 90–150 mm, that of the oranges 60–100 mm, and that of the lemons 45–70 mm.
Assuming normal distributions of the fruit sizes within each class, there will be some misclassified examples if we set hard boundaries at 65 and 95 mm, but the bulk of the fruit will end up in the correct processing line. So far, so good. Our classification scheme will work well as long as the processing plant does not switch suppliers and the supplier keeps delivering produce of the same typical size.
But now assume that, say over the period of one week, the deliveries from the original supplier get gradually replaced by citrus from a different one, with the diameter of 90–150 mm for the grapefruit, 50–75 mm for the oranges, and 70–90 mm for the lemons. By the end of the week, most lemons will end up in the line for oranges and vice versa if we don't retrain our model, using more recent information about the fruit sizes.
Matters get further complicated in a case in which the diameter for all three fruit lies in the range of 70–90 mm and we cannot separate them into three classes solely based on their diameter. In this situation, additional features, such as color or details of their shape, would be required for an accurate classification.
The problem of addressing the concept drift consists in maximizing the classification accuracy at all times while the typical sizes of delivered fruit are changing; this, in turn, leads to the problem of finding the optimal training window size in time at any given time. One way to tackle this problem could be using Gaussian classes as described in Chapter 3 of [27].

Appendix B Predicting daily EURUSD returns

We use directional one-day forecasts for EURUSD to illustrate the dependence of the accuracy of supervised learning forecasts for time series on the training window size. Regardless of limitations of their practical applicability,[5] we demonstrate that the best classification results are generally not obtained for the largest training window size.
Publicly available daily EURUSD close data (reflecting the exchange rate at midnight GMT) for the years 2004–2015 were downloaded from Dukascopy Bank SA [18]. We edited the data by first deleting Saturdays and Sundays. (For some reason, the dates in the downloaded data were shifted by one day, so we had to delete Fridays and Saturdays to omit weekends.) This first preprocessing step resulted in the data set that we call "Dukascopy." In a second preprocessing step, we compared the data on the remaining days with proprietary data obtained through our institutional account

[5] As our results illustrate, a simple directional (up/down) prediction accuracy of more than 50% does not necessarily result in profitable trades: in addition to the impact of transaction costs, it is possible that misclassifications happen more often for forecast instances with large moves than correct classifications, resulting in an overall loss.

with OANDA Corporation [19] and deleted a few days from the data labeled "Dukascopy" to align the two data sets. This resulted in two data sets that each contain 3126 days and that we call "DukascopyAligned" and "OANDAAligned." For the 3126 days, the median absolute difference between DukascopyAligned and OANDAAligned is 9.0×10^{-5} and the mean is 2.2×10^{-4}, which is on the order of one pip (0.0001). The standard deviation of the difference is 4.6×10^{-4} and the mean 2.1×10^{-5}; the largest discrepancies are occurring, not surprisingly, before and after holidays when trading is thin and the data may also reflect the daily close recorded at inconsistent times of the day. We also note that the difference between the data sets is smaller for more recent data. The time series as well as the decadic logarithm of the absolute difference between the two data sets are shown in Figure 4.

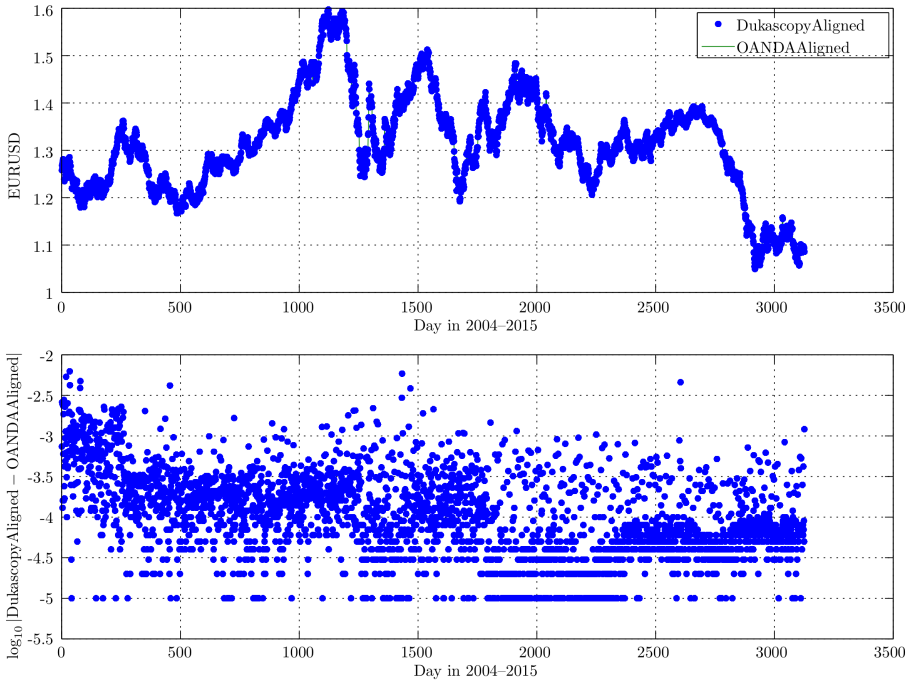


Figure 4: Daily EURUSD close for 2004–2015 for data from Dukascopy Bank SA aligned with OANDA (top), and the decadic logarithm of the absolute difference between the two data sets (bottom). See text for details.

We further divided OANDAAligned into eight five-year periods from 2004–2008 to 2011–2015. For each of the five-year periods of approximately equal lengths, the portion of the five-year set used for training ranged in 5%-steps from 5% to 80% (approximately three months to four years), corresponding to relative training window sizes of $0.05, 0.10, \ldots, 0.80$. The models' performance was evaluated in terms of the accuracy of one-day forecasts on the most recent 20% of the data set (one year); thus, for each of the runs with different training window sizes, the model was tested on the same data. The model was retrained on a daily basis, using the most recent training window for each of the one-day predictions.

Within the scope of the present article and for the purpose of illustrating the relationship between the training window size and the accuracy, the computations were performed with GNU Octave [28], a high-level programming language mostly compatible with MATLAB®[29]. We made ad-hoc

choices for the classifier, using Naïve Bayes [30] from the NaN toolbox [31], as well as for the features (lagged log-differences) and the number of features included. We also obtained results of numerical experiments that included the use of other feature sets (e.g., slopes of moving averages), the use of several feature selection methods, and the use of other classifiers. This included a version of our GNU Octave code that interacts with various tools from scikit-learn [32], a machine learning library for Python. The results differed in details, but lead to conclusions identical to the ones below.

The top portions of Figures 5 (a)–(d) show the up/down classification accuracy for one-day forecasts as a function of the portion of the 2011–2015 five-year data used for training. The cumulative returns obtained for the training window sizes resulting in the best accuracy can be seen in the bottom parts.

Using the daily close values y_d for each of the days d in the data set, we made an ad-hoc choice for the n features given by the logarithmic differences

$$f(d,i) = \ln\, y_d - \ln\, y_{d-i} \quad \text{where} \quad i = 1, \ldots, n\ . \tag{4}$$

For the results presented in Figure 5 (a) and (b), we chose $n = 10$ to include daily close values in the features going back 10 days.

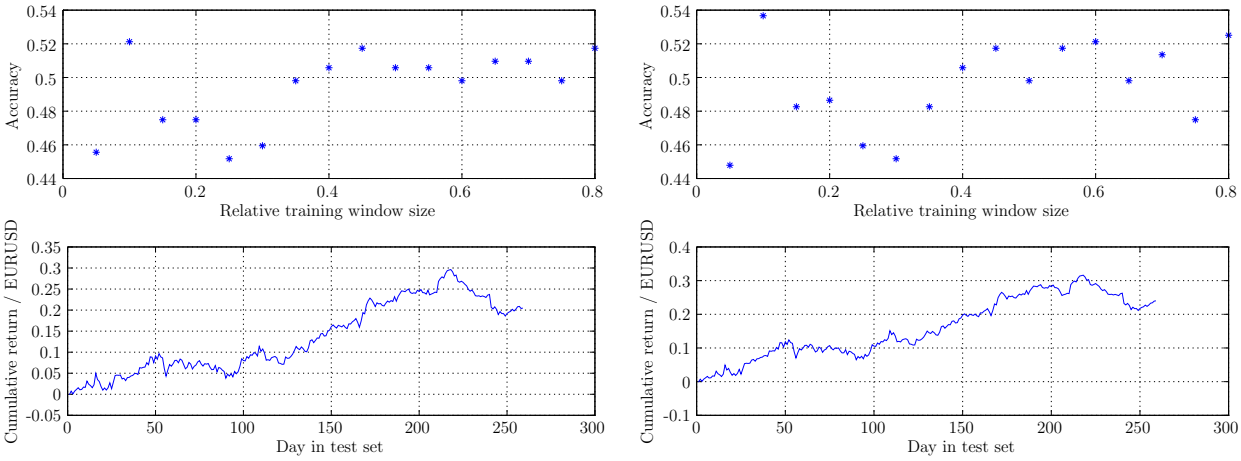
For DukascopyAligned as well as for OANDAAligned, the best classification accuracy is achieved when 10% of the data (six months) are used for training.

In this example, the classification accuracy as well as the cumulative return are slightly better for the OANDAAligned data set than for DukascopyAligned; but, we need to emphasize that we are only showing one specific time interval, and we made an ad-hoc choice for the features as well as the classifier.

A different ad-hoc choice for the feature set was made in (c) and (d): The only difference between the parameters used to produce the results in (a) and (b) in contrast to (c) and (d) is that the latter were computed with $n = 20$. We see that using the DukascopyAligned data set yields a higher accuracy and also a better return than OANDAAligned. For the DukascopyAligned (OANDAAligned) data set, the best classification performance is achieved when 40% (5% or 40%) of the data, i.e., two years (three months or two years), are used for training.

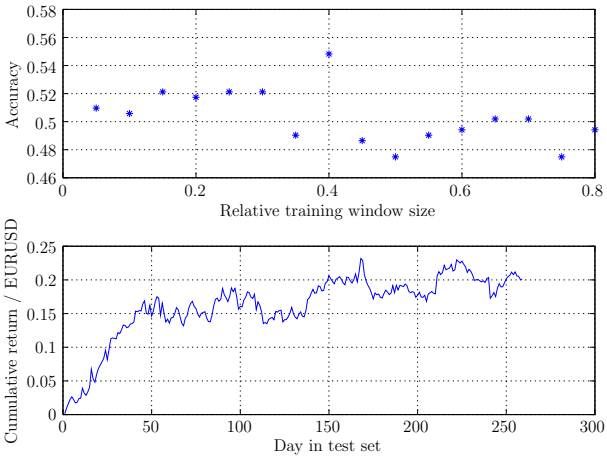
As our more comprehensive exploratory studies (data not shown) suggest, the apparent better performance for one data set compared to another one for certain cases is entirely an artifact of certain parameter choices; results for other parameter choices (e.g., number of features used) and data (as we confirmed by using proprietary data from a third source; data not shown) vary widely. As noted above, the differences between the data (DukascopyAligned vs. OANDAAligned) are on the order of one pip. This is well over one order of magnitude less than the typical daily variations of the exchange rate, and any meaningful predictive method for one-day forecasts would ideally be robust with regard to such small differences between data sets.

Furthermore, our studies have shown that the presence or absence of one single daily data point generally has a significant effect on predictions many days into the future, depending on the features and their lag. This is problematic because it is difficult if not impossible to decide which "unusual" days (e.g., Easter Monday, cf. footnote on page 2) should be omitted from historical data in order to obtain a more "typical" representation of the past for training purposes.

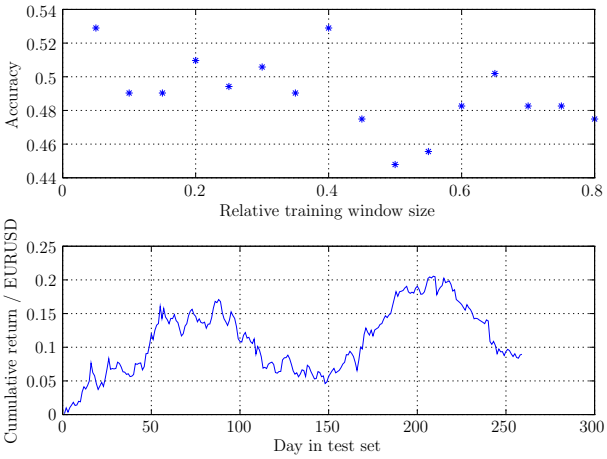


(a) DukascopyAligned

(b) OANDAAligned



(c) DukascopyAligned



(d) OANDAAligned

Figure 5: Classification accuracy (top) as a function of the training window size and cumulative return (bottom) obtained for the window size resulting in the best accuracy. Data from the five-year period 2011–2015 were used. For (a) and (b), we used $n = 10$, and for (c) and (d) $n = 20$. See text for details.

We then conducted a more comprehensive analysis for OANDAAligned and $n = 20$, covering all eight five-year periods from 2004–2008 to 2011–2015. For each of the five-year periods of approximately equal lengths, the portion of the five-year set used for training ranged in 5% steps from 5% to 80% as described above.

For the eight data sets, the left portion of Figure 6 illustrates the distribution of training window sizes for which the best classification accuracy is obtained, and the distribution of values of the corresponding best accuracy is shown in the center. The panel on the right shows the mean accuracy averaged over all eight five-year periods as a function of the relative training window size.

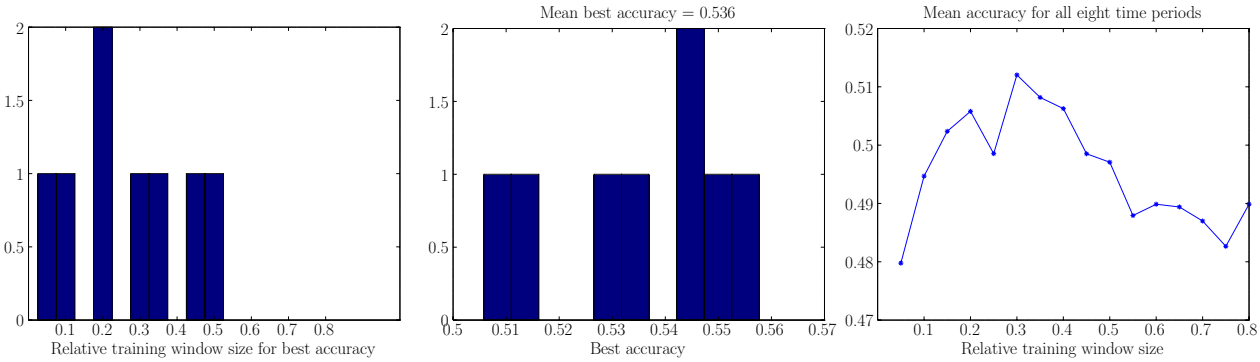
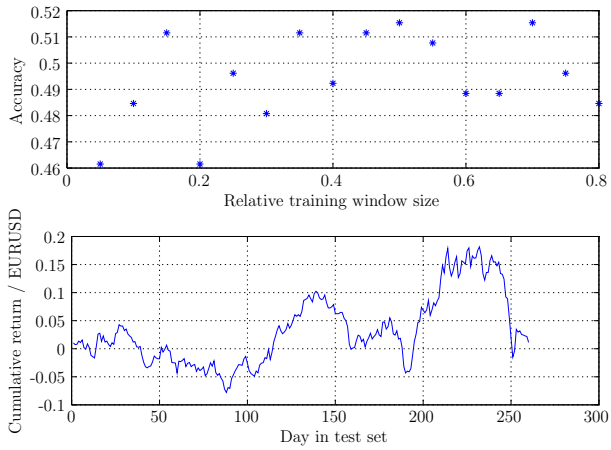


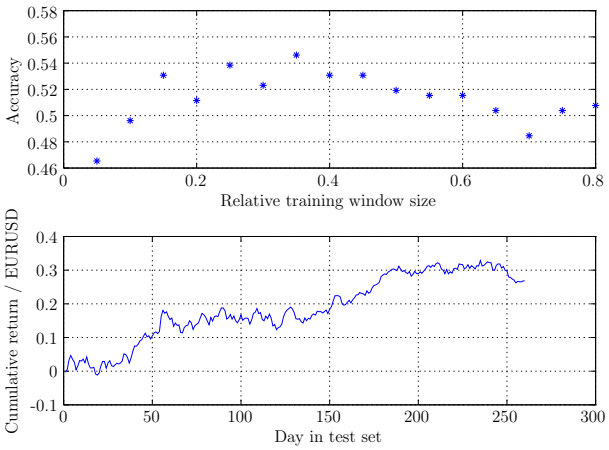
Figure 6: Optimal training window size (left) and corresponding best classification accuracy (center) distributions for the eight five-year periods 2004–2008 to 2011–2015. The panel on the right illustrates the mean accuracy averaged over all eight five-year periods as a function of the relative training window size.

In Figures 7 (a)–(h), we present results analogous to the ones shown in Figure 5 (d) for 2011–2015, but for all eight different five-year periods from 2004-2008 to 2011–2015. The up/down classification accuracy for one-day forecasts as a function of the portion of the five-year data used for training in shown on the top panels, and the cumulative return obtained for the window size resulting in the best accuracy is illustrated on the bottom panels. In Figure 7 (c), e.g., the cumulative return is shown for a relative training window size of 0.10; i.e., approximately six months' worth of data were used to train the models used for each of the one-day forecasts in the test set.

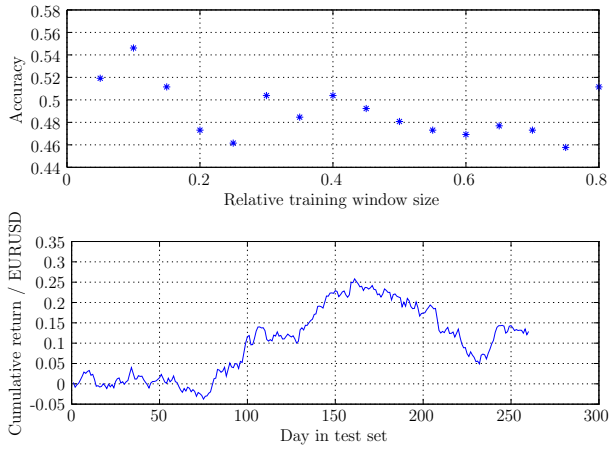
While the present study is by no means comprehensive and the approach has not been optimized with regard to trading profit, our study demonstrates that the best results are not generally obtained for the largest training window size: as shown on the left panel of Figure 6 for the eight five-year periods we evaluated, we obtained the best classification accuracies for relative training window sizes between 0.05 and 0.50. And, while the optimal choice depends on the specific time period considered, the right panel of Figure 6 shows that the mean accuracy, when averaged over all eight time periods, peaks at a relative training window size of 0.3 where it assumes a value of 0.512. This maximum is smaller than the value of 0.536 shown on the center panel, because the latter was computed by averaging the best results that were obtained for the eight periods individually and, as shown on the left panel, for different training window sizes.



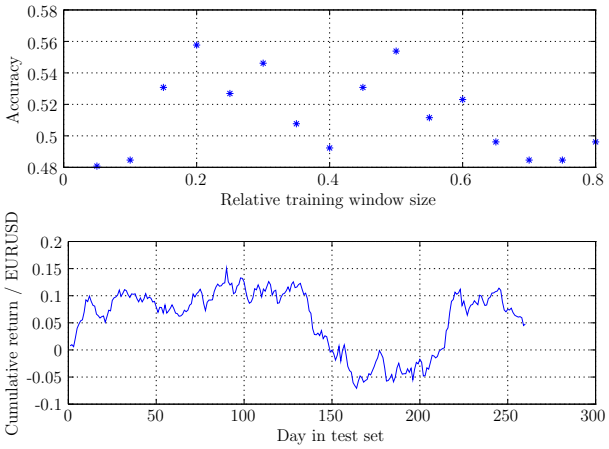
(a) 2004–2008



(b) 2005–2009

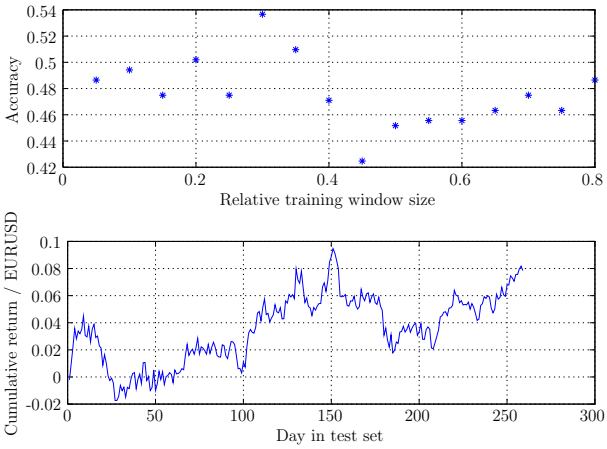


(c) 2006–2010

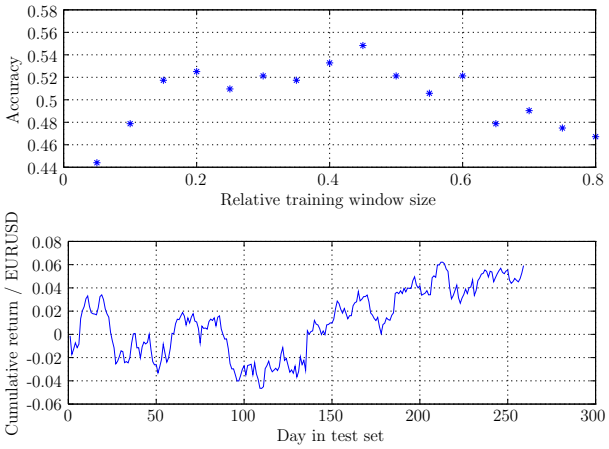


(d) 2007–2011

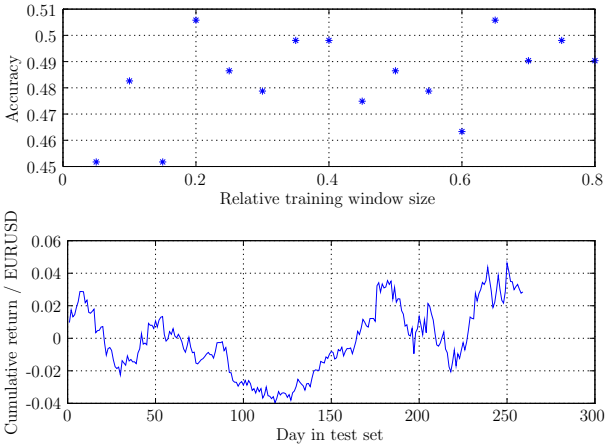
Figure 7: Classification accuracy (top) as a function of the relative training window size and cumulative return (bottom) obtained for the window size resulting in the best accuracy. Four five-year periods from 2004 to 2011 are shown on this page (continued on the next page).



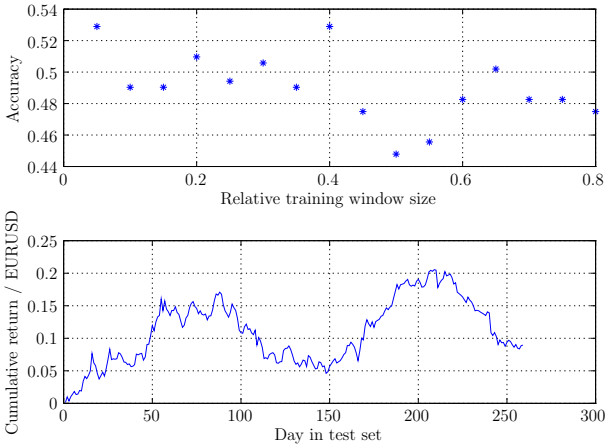
(e) 2008–2012



(f) 2009–2013



(g) 2010–2014



(h) 2011–2015

Figure 7: Classification accuracy (top) as a function of the relative training window size and cumulative return (bottom) obtained for the window size resulting in the best accuracy. Four five-year periods from 2008 to 2015 are shown on this page (continued from the previous page).

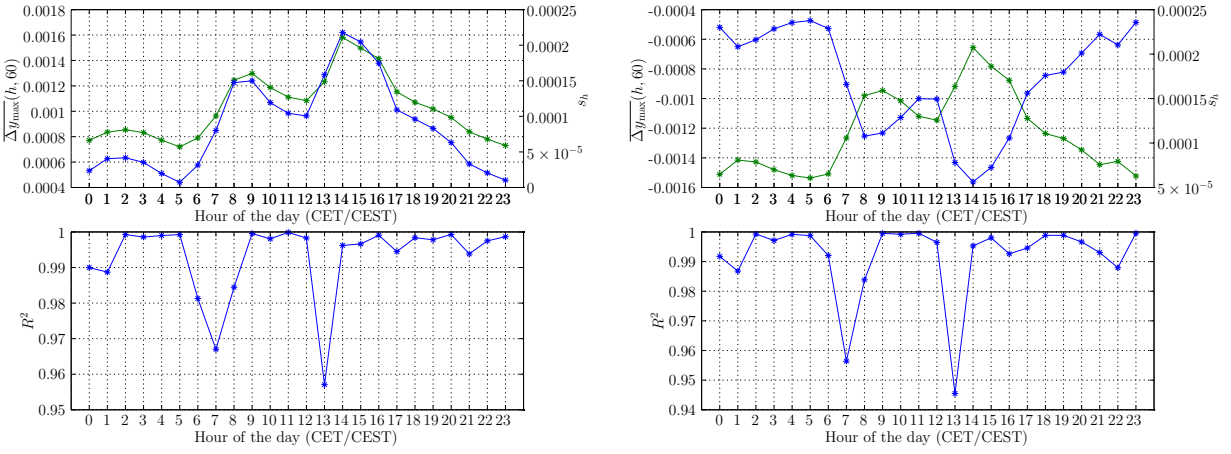
Appendix C A random walk model and its limitations

In this section, we model the behavior of EURUSD as a stochastic process to present a plausible argument for the dependence of $\overline{\Delta y_{\max}}$ on k that is shown in Figure 1.
For a given hour h, we model the currency moves as a Gaussian random walk with normally distributed random changes with variance σ_h^2 and mean zero that are uniformly separated by time intervals of length δt. (For this purpose, we could take $\delta t = 1/f_{\text{tick}}$ with the tick frequency f_{tick} that we assume to be constant.) Then the leading term for the expectation value of the maximum of the walk after n steps will be given by $\Delta y_{\max_n} \sim \sqrt{n}$ [33, 34, 35]. Per construction of our model, we have 1 minute $= \alpha\,\delta t$ and therefore $n = \alpha\, k$, with some positive constant α. Thus, if our model assumptions are valid, we should expect

$$\overline{\Delta y_{\max}}(h,k) \approx s_h\sqrt{k} \quad \text{and} \tag{5}$$

$$\overline{\Delta y_{\min}}(h,k) \approx -s_h\sqrt{k} \tag{6}$$

with a positive constant s_h that depends on the hour of the day as well as on the volatility during that time of the day.
The following numerical results are based on the same one-minute data for EURUSD from July 2012 until March 2016 that we used in Section 2.1 (H212, . . . , H116). For H212–H116, we performed non-linear fits of functions $\pm s_h\sqrt{k}$ to the data to facilitate the comparison with fits of functions that include additional terms [34, 35] (data not shown); the results shown in Figures 8, 9, and 11 are consistent with those from linear fits that we performed to $\overline{\Delta y_{\max}}^2$ and $\overline{\Delta y_{\min}}^2$. Figure 8 (a) shows the mean of the 60-minute maximum increase, $\overline{\Delta y_{\max}}$, and part (b) shows the mean decrease, $\overline{\Delta y_{\min}}$, together with the coefficients s_h from the least squares fits as well as the coefficients of determination, R^2. The behavior of s_h mirrors the volatility as we had previously defined it in Equations (1)–(3).



(a) $\overline{\Delta y_{\max}}(h,60)$ (b) $\overline{\Delta y_{\min}}(h,60)$

Figure 8: The top panels show the extrema of changes of EURUSD over 60-minute intervals, $\overline{\Delta y_{\max}}(h,60)$ and $\overline{\Delta y_{\min}}(h,60)$ (blue, left scales), and the least squares fit coefficients s_h (green, right scales). The coefficients of determination R^2 are shown in the lower panels. See text for details.

The least squares coefficients s_h from the fits to $\overline{\Delta y_{\max}}$ and $\overline{\Delta y_{\min}}$ are not identical as they would be for a symmetric Gaussian random walk; but, they only differ by at most a few percent. Furthermore, the high R^2 values indicate that the $\sim \pm\sqrt{k}$ approximations describe the extrema rather well. This appears astonishing, especially in light of several crude assumptions that we made and that are by no means accurate representations of reality. (For example, we made the assumptions that, for a given hour of the day, the tick frequency as well as the volatility are constant and that the changes are normally distributed.)

It is notable, however, that there are two dips of R^2 at around the hours 7 and 13. To gain insight into one possible reason for the deviations from the model $\sim \pm\sqrt{k}$, it is instructive to realize that both hours coincide with times of increasing volatility:

In Figure 9, we can see both $\overline{\Delta y_{\max}}$ and $\overline{\Delta y_{\min}}$ (blue) in comparison with the fitted functions (green) for different times of the day. For the hours 5 and 9, shown in (a) and (b), the approximation by the function $\sim \pm\sqrt{k}$ is better than for the hours 7 and 13 that are shown in (c) and (d). For $\overline{\Delta y_{\max}}$ in part (c), for example, we see that the slope is decreasing more slowly than it would for a square root, and for $\overline{\Delta y_{\min}}$ we observe that the slope is increasing more slowly than that of a negative square root. This behavior is plausible, because the hours 7 and 13 fall into periods of increasing volatility; therefore, for larger values of k, the "effective mean volatility" is larger. As a consequence, a larger proportionality constant would be required for the square root of k to get a good fit for larger values of k. In contrast, the volatility remains roughly unchanged following the hours 5 and 9; hence, a better fit can be accomplished with a constant s_h for those times.

But there is another observation that we make when we compare $\overline{\Delta y_{\max}}$ and $\overline{\Delta y_{\min}}$: while the positive and negative branches are approximately the same for the hours 5 and 9, we see that, for the hours 7 and 13, $\overline{\Delta y_{\min}}$ is larger in magnitude than $\overline{\Delta y_{\max}}$. Is it possible that there is a negative bias during the hours 7 and 13? It turns out that the answer is yes; as we discuss in Section 2.2, a negative cumulative average return for EURUSD during the European session has been reported in the literature [25]. (See the bars on the top left graph of Figure 1 or the line on the top left graph of Figure 3 therein.) But, based on the results published in [25], neither one of the hours 7 and 13 (hours 01–02 and 07–08 New York time in [25]) yields a return that is significantly different from zero, and, in particular, the return is not negative; thus, it seems that we cannot utilize the depreciation as a plausible argument for the differences between $\overline{\Delta y_{\max}}$ and $\overline{\Delta y_{\min}}$.

However, for the period from July 2012 to March 2016 that we included in our data, we have been able to use a stochastic approach within the framework of our production environment to identify both hours in question as times of significant decline of EURUSD. As an alternative explanation, we now provide a reasoning based on one-minute close data.

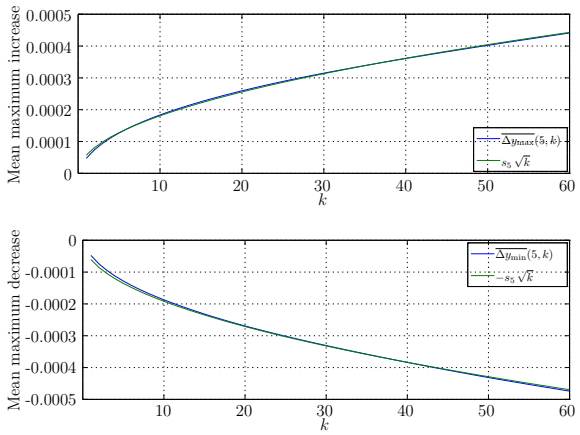
We define the mean hourly return in a manner analogous to the definition of the mean volatility in Equation (3) in Section 2.1: with the k-minute return at time t_0 defined as

$$\Delta y \equiv y_k - y_0 \;\text{ where }\; y_k = y(t_k) \;\text{ with }\; t_k = t_0 + k\,\text{minutes}\,, \tag{7}$$

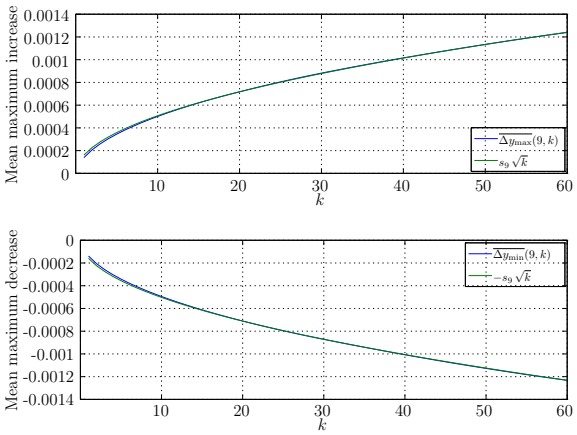
the definition of the mean hourly return for the hour $h = 10$, for example, is

$$\overline{\Delta y_{\text{hourly}}}(h = 10) \equiv \text{mean}\{\Delta y \mid t_0 = 10{:}01, 10{:}02, \ldots, 11{:}00 \;\text{ and }\; k = 60\}\,. \tag{8}$$

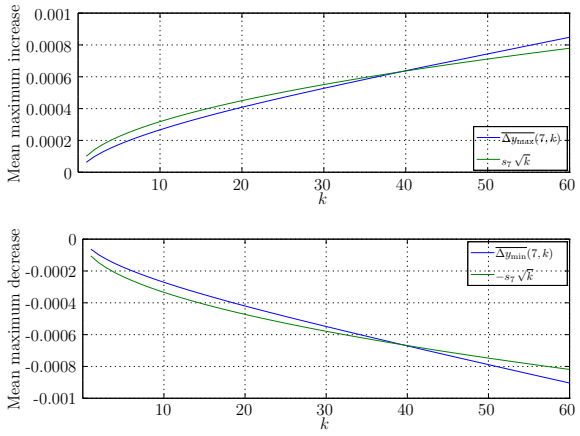
Analogously to the mean maximum increase and decrease, we are using 60 different values per hour to compute the mean hourly return. (This is different from the methodology used in [25], which is based on a comparison of midquote prices at the beginning and end of each hour.) We then



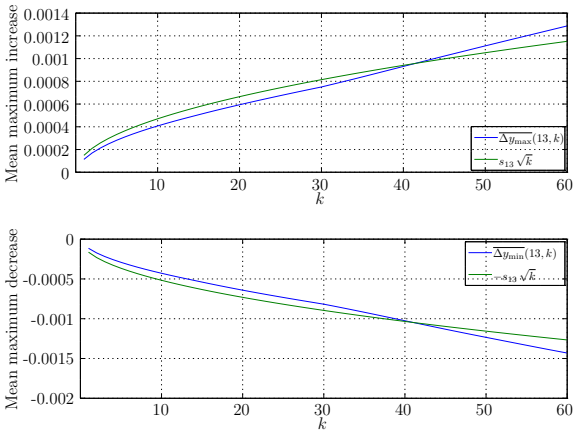
(a) $h = 5$



(b) $h = 9$



(c) $h = 7$



(d) $h = 13$

Figure 9: $\overline{\Delta y_{\max}}$ and $\overline{\Delta y_{\min}}$ (blue) in comparison with the fitted functions $\pm s_h\sqrt{k}$ (green) for different hours h of the day. See text for details.

use the mean hourly return to compute the cumulative mean daily return, which we annualize by multiplying it with 250 (the approximate number of trading days per year).

Figure 10 shows the annualized cumulative mean daily return. We note that the overall variation is on the same order of magnitude as the results presented in [25] for 1997–2007, and we can confirm the mean depreciation of EURUSD during the European session that can be utilized for a trading strategy that is as least marginally profitable after accounting for transaction costs. (We present the return while the log return is shown in [25]; but, since the return is small, the difference is negligible for this comparison.) But we would like to point out that there are significant differences between our results and the ones reported in [25]. In particular, the decline during the European session, according to our analysis, begins earlier and ends later: mean hourly returns (as defined

above) are negative from hour 7 until hour 13. If we take the mean for a smaller value of k, the period of depreciation extends even further into the North American session (data not shown).

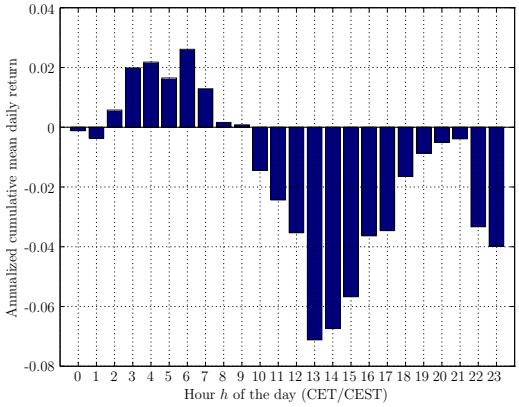
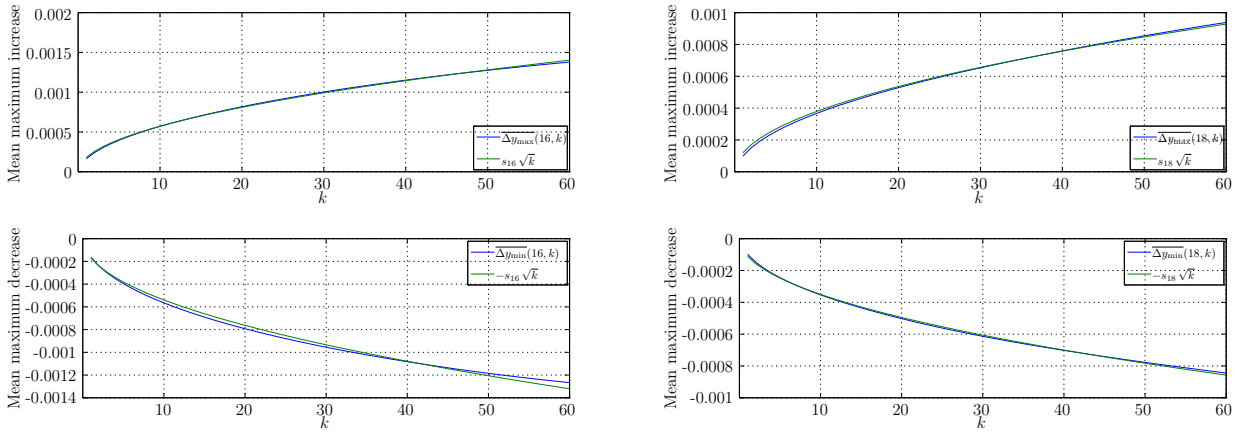


Figure 10: Annualized cumulative mean daily return for EURUSD from July 2012 until March 2016. Like in the results presented in Figures 1 and 2, the bars at the end of the day after the close of the North American markets should not be taken too seriously, because we did not exclude weekend gaps and early market closing days.

This time pattern is consistent with conclusions from statistical significance tests associated with our production code that we mentioned in Section 2.1 and also above in this section.
Figure 11 shows $\overline{\Delta y_{\max}}$ and $\overline{\Delta y_{\min}}$ in comparison with the fitted functions for $h = 16$ and $h = 18$. The picture is the reverse of the situation for $h = 7$ and $h = 13$ above, albeit to a lesser extent: during $h = 16$, the mean return is positive, and the volatility is decreasing somewhat. This is also the case for $h = 18$, but to an even lesser extent and resulting in a better fit of $\pm s_{18}\sqrt{k}$ to the data.



(a) $h = 16$ (b) $h = 18$

Figure 11: $\overline{\Delta y_{\max}}$ and $\overline{\Delta y_{\min}}$ (blue) in comparison with the fitted functions (green) $\pm\, s_h\, \sqrt{k}$ for $h = 16$ and $h = 18$. See text for details.

References

[1] B. E. Boser et al., Proceedings of the Fifth Annual Workshop on Computational Learning Theory, 144–152. (1992)

[2] C. M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, New York. (1995)

[3] B. Krollner et al., European Symposium on Artificial Neural Networks: Computational and Machine Learning. Bruges, Belgium. (2010)

[4] T. Imam, Progress in Intelligent Computing and Applications 1(1), 1–15. (2012)

[5] M. A. H. Dempster et al., IEEE Transactions on Neural Networks 12(4), 744–743. (2001)

[6] J. Kamruzzaman et al., Proceedings of the Third IEEE International Conference on Data Mining, ICDM 2003, 557–560. (2003)

[7] K. Kim, Neurocomputing 55, 307–319. (2003)

[8] R. Choudhry and K. Garg, International Journal of Computer, Electrical, Automation, Control and Information Engineering 2(3), 689–692. (2008)

[9] C. Ullrich et al., in Simeon J. Simoff, Graham J. Williams, John Galloway and Inna Kolyshkina, Proceedings of the 4th Australasian Data Mining Conference, Sydney, Australia, 221–240. (2005)

[10] C. Ullrich, Forecasting and Hedging in the Foreign Exchange Markets, Lecture Notes in Economics and Mathematical Systems 623, Springer. (2009)

[11] M. Alamili, M. Sc. Thesis, Delft University of Technology. (2011)

[12] A. A. Baasher and M. W. Fakhr, Proceedings of the 11th WSEAS international conference on Applied computer science, Penang, Malaysia, WSEAS, 41–47. (2011)

[13] H. Talebi et al., Procedia Computer Science 29, 2065–2075. (2014)

[14] I. Žliobaitė, arXiv preprint arXiv:1010.4784v1. (2010)

[15] J. Gama et al., ACM Computing Surveys 46(4). (2014)

[16] S. Walczak, Journal of Management Information Systems 17(4), 203–222. (2001)

[17] M. D. Rechenthin, Ph. D. thesis, University of Iowa. (2014)

[18] Dukascopy Bank SA, Geneva, Switzerland, Historical Data Feed, `https://www.dukascopy.com/swiss/english/marketwatch/historical/`

[19] OANDA Corporation, Toronto, Canada, `http://www.oanda.com/`

[20] Perl DateTime Module, `http://search.cpan.org/dist/DateTime/`

[21] CBOE Volatility Index®,`http://www.cboe.com/micro/vix/vixintro.aspx`

[22] T. Ito and Y. Hashimoto, NBER Working Paper No. 12413. (2006)

[23] S. B. Achelis, Technical Analysis from A to Z, Second Edition, McGraw-Hill. (1995)

[24] Cornett et al., Journal of Banking and Finance 19, 843–869. (1995)

[25] F. Breedon and A. Ranaldo, Queen Mary, University of London, School of Economics and Finance Working Paper No. 694. (2012)

[26] OECD, International Standards for Fruit and Vegetables (Citrus), http://www.oecd.org/agriculture/code/43579800.pdf

[27] I. Žliobaitė, Ph. D. thesis, Vilnius University. (2010)

[28] J. W. Eaton et al., GNU Octave version 3.8.1 manual: a high-level interactive language for numerical computations. CreateSpace Independent Publishing Platform. ISBN 1441413006. (2014)

[29] MATLAB®, The MathWorks Inc., Natick, Massachusetts, http://www.mathworks.com

[30] H. Zhang, International Journal of Pattern Recognition and Artificial Intelligence 19(2), 183–198. (2005)

[31] A. Schloegl, The NaN-toolbox: A statistics and machine learning toolbox for Octave and MATLAB®, http://pub.ist.ac.at/%7Eschloegl/matlab/NaN/

[32] F. Pedregosa et al., Journal of Machine Learning Research 12, 2825–2830. (2011)

[33] E. W. Weisstein, Random Walk–1-Dimensional, from MathWorld – A Wolfram Web Resource. (2002)

[34] A. Comtet and S. N. Majumdar, arXiv preprint arXiv:cond-mat/0506195v2. (2005)

[35] E. G. Coffman et al., Probability in Engineering and Informational Sciences 12, 373–386. (1998)