Short-Term Euro-Dollar Exchange Rate Forecasting Using Regression Models

Alireza Taghi[1,2], Dr. Achraf Cohen[1]

(1) Department of Mathematics and Statistics, University of West Florida, USA

(2) Department of Accounting and Finance, University of West Florida, USA

INTRODUCTION

An exchange rate is the price of one country's currency expressed in terms of some other currency (Adeniran, Yusuf, & Adeyemi, 2014). One of the critical reasons for forecasting the exchange rate is international trade, especially multinational companies in import-export businesses. Moreover, on a larger scale, a change in a country's exchange rate could affect the economic growth of both importing and exporting countries (Aliyu, 2009). Also, individuals can invest in various exchange rates in the forex market and gain capital or protect their monetary value against changes in the currency value of their native currency. The forex market is the biggest trading market in the world in which both individual and institutional traders are investing and trading in the market.

The forex market is known for its high risk and volatility, which comes with significant profit and loss. However, some studies showed that not only retail traders but also individual traders were able to make a profit (Abbey & Doukas, 2015). Given the new tools and models used by institutional traders, such as high-frequency algorithmic trading models, it is more challenging than years ago to profit the old-fashion way using only fundamental and technical analysts. Therefore, it is expected that more individual traders will start using automated and algorithmic trading in the future.

There has been discussion that exchange rates are random and not predictable, well-known as a random walk theory. Nevertheless, as statistics, machine learning, and computing unfold and quickly evolve, we now have more tools to evaluate the models that were not possible a decade ago. For instance, Colombo and Pelagatti (2020) work shows that some statistical learning methods outperformed the result of random walk theory in forecasting except for a short period (1-2 months).

Furthermore, another study using ensemble techniques, which combine linear regression with other methods to elevate the ultimate model performance, was proposed in (Sharma, Hota, & Handa, 2017). These published articles suggest that more research is being done to investigate the exchange rate forecasting problem using machine learning tools and algorithms.

AIM

In this project, our goal is to investigate mathematical and statistical models to forecast the short-term exchange rate. Notably, we are considering the 15-minutes time frame Euro-Dollar (EUR/USD) currency pair as the object of the project.

There are several different major currency pair such as EUR/USD USD/JPY, GBP/USD, USD/CHF, AUD/USD, and USD/CAD.

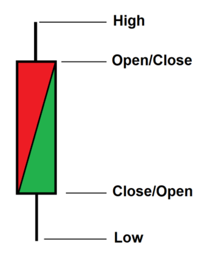
We chose EUR/USD since it is the most traded currency pair in the market; however, most of our work can be applied to other currency pair with some modifications.

In addition, we will also focus on polynomial regression models, which we hypothesized to be a better fit given the non-linear nature of the data.

DATA

We obtain the data from one of the forex brokers, OANDA, through the MetaQuotes software. Every 15 minutes an observation (candle) is collected, and it is made of four key variables as follow:

- **Open** (price at beginning of each 15-min)
- **High** (highest price during the 15-min)
- **Low** (lowest price during the 15-min)
- **Close** (price at the end of each 15-min)



As one can expect in order to predict the next candle we need to predict Open, High, Low, and Close. We used 24,900 data point from January to December of 2020. Also notice that the color of candle is different depending on being increasing or decreasing. Decreasing candle is set to be red ($\text{Open} > \text{Close}$), and increasing candle is green ($\text{Close} > \text{Open}$).

METHOD

Multiple Regression: The general format of multiple regression models is shown below. we can see Y is the dependent variable, x's ($x_1, x_2, \dots, x_k$) are independent variables, β's are coefficients, and the ε is the error term.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k + \varepsilon$$

In the case of multivariable multiple regression, there would be more than one dependent variable: Y's ($Y_1, Y_2, \dots, Y_k$).

T^2 Hotelling Statistic: Hotelling's T-squared statistic is the generalized version of student t-statistic, which will enable us to handle comparison in a multivariate data (Hotelling, 1931). Consider $\mathbf{X_i} = (X_1, X_2, X_3, \dots, X_p)_i$ a p-variate data. The T-squared statistic for an observation i:

$$T_i^2 = (X_i - \bar{X})'\Sigma^{-1}(X_i - \bar{X})$$

Where $\bar{X}$ is a mean vector and Σ^{-1} is the inverse of the matrix of covariances variances.

Kolmogorov–Smirnov Test: It is a non-parametric test to compare empirical distributions of two independent groups. It is "defines the largest absolute difference between the two cumulative distribution functions as a measure of disagreement" (Lopes et al., 1970)

MODELS & PROCEDURES:

Multivariate Multiple Linear Regression Model (MMR):

$$[\text{Open}_t, \text{High}_t, \text{Low}_t, \text{Close}_t] \sim [\text{Open}_{t-1} + \text{High}_{t-1} + \text{Low}_{t-1} + \text{Close}_{t-1} + \text{vol}_{t-1} + \text{weekDay}_{t-1} + \text{hours}_{t-1} + \text{market}_{t-1} + \text{up}_{t-1}]$$

Multivariate Multiple Polynomial Regression Model (Poly):

$$[\text{Open}_t, \text{High}_t, \text{Low}_t, \text{Close}_t] \sim [\text{Open}_{t-1} + \text{Open}^2_{t-1} + \text{Open}^3_{t-1} + \text{High}_{t-1} + \text{High}^2_{t-1} + \text{High}^3_{t-1} + \text{Low}_{t-1} + \text{Low}^2_{t-1} + \text{Low}^3_{t-1} + \text{Close}_{t-1} + \text{Close}^2_{t-1} + \text{Close}^3_{t-1} + \text{vol}^3_{t-1} + \text{weekDay}_{t-1} + \text{hours}_{t-1} + \text{market}_{t-1} + \text{up}_{t-1}]$$

Summary of Procedures:

We fitted two regression models: a multivariate multiple linear model (MMR) and multivariate multiple polynomial model (Poly) in order to forecast the rate of Open, High, Low, and Close which a sample of the results from the May 20, 2020 demonstrated in figure 1.

We trained the models on the lagged 15 minutes data of the previous trading week. Our independent variables include Open, High, Low, Close, Volume (vol), Day of the Week (weekDay i.e. Mon, Tus, …), hours, market session, and increasing/decreasing factor (up).

We use the T^2 Hotelling's statistic to analyze the errors from each model. T^2 statistic will allow us to measure the performance of each model (MMR and Poly). We calculate the prediction error as the difference between the predicted and the actual values Open, High, Low, Close in the unit of pips. Four different error values were obtained and compared to a vector of four zeros (perfect prediction).

A Sample of the results: each box represent a timestamp which includes MMR, Poly, and Actual candles

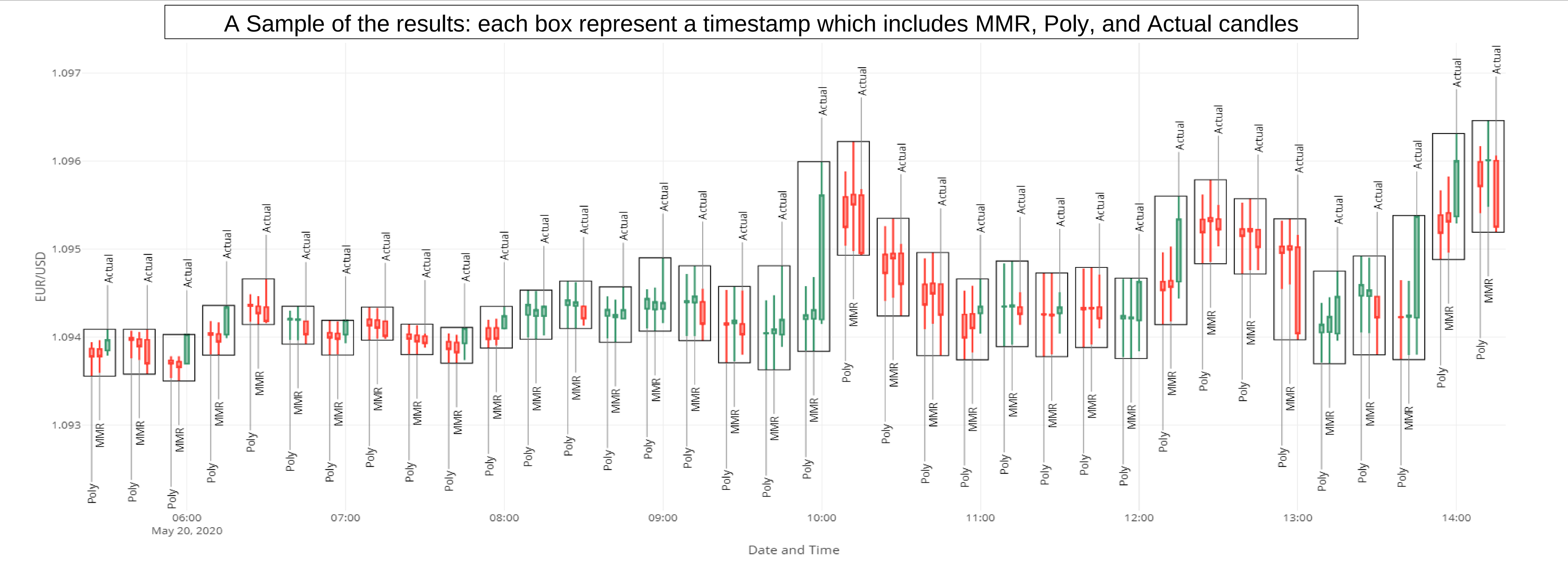


Figure 1

We would then compute the T^2 Hotelling statistics for each [Open, High, Low, Close] prediction. A smaller T^2 indicates accurate predictions.

After obtaining the T^2 Hotelling statistics for the two models, we compare the two distributions of T^2 Hotelling statistics (MMR vs. Poly) using the Kolmogorov-Smirnov test, which the density plot is demonstrated below in the figure 2.

Density Plot of T-squared for MMR and Poly Models

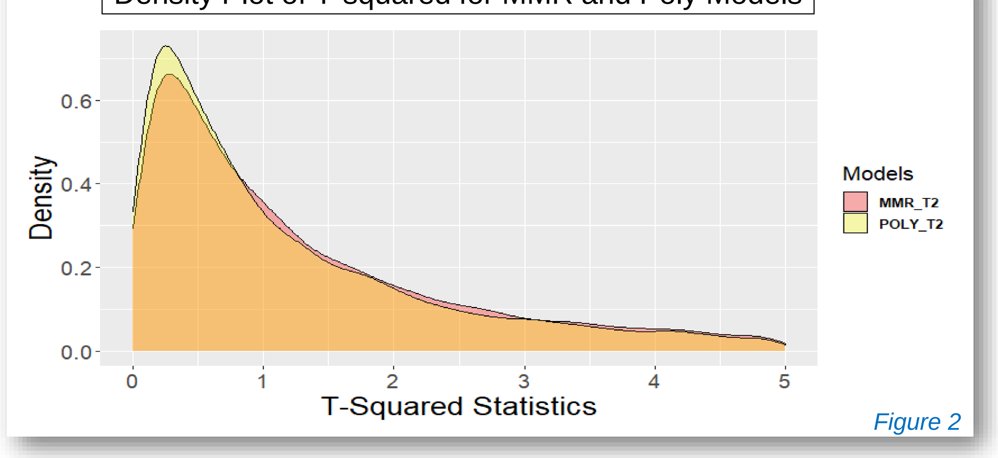


Figure 2

RESULT & DISCUSSION

We obtained a highly significant p-value <0.0001 from the Kolmogorov-Smirnov test, and we conclude that overall, the values of T^2 statistic based-error for the polynomial model are smaller than those from the linear model. Therefore, as we hypothesized, we can conclude that the overall result of model Poly is better and more accurate than the MMR.

Although the T^2 Hotelling statistic enables us to compare models to each other, it is very challenging, if all possible, to interpret the T^2 directly in terms of the probability of success for trading purposes which we would be investigating in our future studies.

REFERENCES

Abbey, B. S., & Doukas, J. A. (2015). Do individual currency traders make money? Journal of International Money and Finance, 56, 158-177. doi:10.1016/j.jimonfin.2014.10.003

Adeniran, J., Yusuf, S., & Adeyemi, O. A. (2014). The Impact of Exchange Rate Fluctuation on the Nigerian Economic Growth: An Empirical Investigation. International Journal of Academic Research in Business and Social Sciences, 4(8). doi:10.6007/ijarbss/v4-i8/1091

Aliyu, S. U. (2009). Impact of Oil Price Shock and Exchange Rate Volatility on Economic Growth in Nigeria: An Empirical Investigation. SSRN Electronic Journal. doi:10.2139/ssrn.1346418

Colombo, E., & Pelagatti, M. (2020). Statistical learning and exchange rate forecasting. International Journal of Forecasting, 36(4), 1260-1289. doi:10.1016/j.ijforecast.2019.12.007.

Hotelling, H. (1931). The generalization of student's ratio. *The Annals of Mathematical Statistics*, *2*(3), 360–378. https://doi.org/10.1214/aoms/1177732979

Lopes, R. H. C., Reid, I. D., & Hobson, P. R. (1970, January 1). *The two-dimensional Kolmogorov-smirnov test.* Brunel University Research Archive: Home. https://bura.brunel.ac.uk/handle/2438/1166.

Sharma, D. K., Hota, H. K., & Handa, R. K. (2017). PREDICTION OF FOREIGN EXCHANGE RATE USING REGRESSION TECHNIQUES. Review of Business and Technology Research, 14(1), 29-33. Retrieved from http://mtmi.us/rbtr/dec/dec2017/18-Handa_Shrivas_Hota-pp111-116.pdf.