An Overview of Gaussian process Regression for Volatility Forecasting

1st Bingqing Liu
Department of Engineering
University of Oxford
Oxford, United Kingdom
bingqingliudoris@gmail.com

1st Ivan Kiskin
Department of Engineering
University of Oxford
Oxford, United Kingdom
ikiskin@robots.ox.ac.uk

3rd Stephen Roberts
Department of Engineering
University of Oxford
Oxford, United Kingdom
sjrob@robots.ox.ac.uk

***Abstract*—Forecasting financial time series using trading data has held the attention of academics and practitioners due to the complexity of the financial system and the profit it can generate for investors. Although investors aim to achieve a consistent attainment of returns, it is important for investors to understand the concept and measurement of volatility. A higher volatility indicates a wider potential range of future returns. But with higher potential returns, comes a higher potential risk. With Gaussian Processes (GPs) having been shown great potential in this field, this paper explores the application of GP regression in forecasting the volatility of foreign exchange returns. This paper builds on the existing literature by applying Gaussian processes for time series forecasting in new ways, namely the multivariate non-coregionalised and coregionalised GP. We show that a multivariate GP can match the accuracy of predictions of a univariate GP, with the added benefit of lower predictive uncertainty due to the incorporation of extra information. Furthermore, we give insight into the relative strengths of the GP methods with recommendations to the practitioner.**

***Index Terms*—Volatility, Forex, Gaussian Process, Regression, Multivariate**

I. Introduction

A financial market is a place where buyers and sellers buy and sell assets. The global financial market is heavily automated and data-driven due to the large volume of trades. Investors use statistical models that can capture market behaviour, trends, and patterns using huge quantities of data and various techniques [1], [2]. Volatility forecasting is an important task in financial markets and several approaches have been used to study and estimate the performance of different volatility models. Volatility is defined as a statistical measure of dispersion of returns of a financial asset. It is commonly measured using the standard deviation of asset time series data or the absolute returns as a proxy measure. Commonly, the higher the volatility, the riskier the asset. Therefore, forecasting volatility provides an insight into financial risk management [3].

Our approach to predicting financial volatility in this paper uses Gaussian process (GP) regression [4]. This is a Bayesian non-parametric kernel-based probabilistic model, commonly used in time-series modelling [5], [6]. We compare and contrast three implementations of GPs, namely a univariate GP, a multivariate non-coregionalised GP, and a multivariate coregionalised GP. With a rolling-window algorithm, we show that each GP is able to forecast the volatility, and we note the differences betwen the approaches. We furthermore show how the error depends on the prediction horizon employed.

The remainder of the paper is structured as follows. Section II gives an overview of the literature surrounding appropriate financial time-series modelling. Section III defines the mathematical quantities and techniques used in the paper. We describe different volatility forecasting models, including univariate and multivariate GPs. Section IV describes the method of processing and analysing data, as well as outlines the approach for contrasting the various models. Section V discusses the results of the forecasting approaches. Finally, Section VI concludes.

II. Related Work

A widely used model to forecast volatility is the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model [7]. The GARCH model is a generalization of the ARCH (autoregressive) model. Other modified GARCH models such as the exponential GARCH model are also commonly used to overcome the limitation of GARCH. In [8], the authors describe and evaluate various popular time series volatility models that use the historical information set to formulate volatility forecasts. The paper focuses on ARCH class conditional volatility models, stochastic volatility models, historical volatility models that include random walks, historical averages of squared returns or absolute returns, and options-based volatility forecasts. By carefully reviewing the methodologies and findings in 93 papers, the authors conclude that the options-based volatility provides the best forecasting, and the historical volatility models slightly outperform the GARCH models.

Recently there has been a shift of focus from parametric models, such as in [8], to semi-parametric and non-parametric models to forecast volatility [9]. Bayesian non-parametric models attracted more attention and there have been a few published works that explore the application of Gaussian process regression on volatility forecasting. In [10], [11], the authors compared the Gaussian process model to the GARCH model and conclude that The GP model outperforms standard stochastic volatility and GARCH models in one-month-ahead out-of-sample volatility forecasting. We build on this by providing alternative implementations of GPs, and provide insights into their relative caveats and merits. Furthermore, the authors in [12] use GP regression to propose an approach for predicting volatility of financial returns by forecasting the envelopes of the time series. This concept has been extended by applying multivariate Gaussian process regression for Portfolio Risk Modeling [13]. Due to the promise shown by GPs, we choose to compare a range of GP approaches to illustrate and contrast new ways of working with GP models for volatility forecasting.

III. Mathematical definitions

This section gives a mathematical overview of the definitions of volatility and related quantities used in this paper, as well as the different Gaussian process models.

A. Forex market

The foreign exchange market (Forex market) is a global over-the-counter market for trading currencies. It is the most liquid financial market in the world and the market participants include commercial companies, central banks and other financial institutions. All the calculations and inference of the volatilities in this paper are based on Foreign Exchange market data of EUR/CHF, and EUR/USD, where EUR/CHF stands for the amount of CHF that an EUR can buy. The data are obtained at 30-minute intervals over a two-year period. As an example, Figure 1 shows the currency exchange rate of EUR/CHF over a ten-year period.

Fig. 1: There have been periods where the variance in exchange rate is high (i.e. high volatility) and periods where the variance in exchange rate is low.

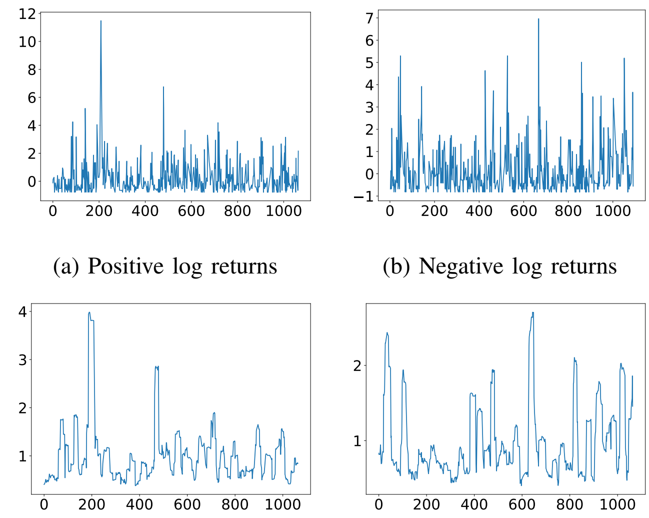
B. Forex price return

Financial returns describe the forward difference of a time series. The benefit of using returns, versus prices, is that returns can measure all variables in a comparable metric and enable analytic relationships amongst different variables.

We use the geometric returns (logarithmic return) in this paper. For the Eurodollar, the geometric returns r_t at time t is:

$$r_t = \log(\mathrm{EUR/USD}_t) - \log(\mathrm{EUR/USD}_{t-1}). \tag{1}$$

Logarithmic returns reduce the data skewness and by taking the difference, the data is de-trended, removing non-stationarity of the data.



(a) Positive log returns

(b) Negative log returns

(c) Positive volatility

(d) Negative volatility

Fig. 2: Positive and negative log returns in (a), and (b) respectively. Corresponding short-term realised volatility in (c) and (d). All computations are made on EUR/USD data according to Equations 1 and 2.

The first step is to separate positive returns from negative returns due to the asymmetry property of volatility. As Forex price is always one-way quoted, the negative returns are known to have a great impact on volatility [11], [12]. We show the separation of positive and negative log returns in Figure 2 (a) and (b).

C. Volatility computation

It is difficult to directly forecast short-term foreign exchange asset returns due to extreme high frequency variation. To combat this issue, we forecast the short-term realised volatility instead. We define short-term realised volatility as the return of a short window in $L2$ norm:

$$V_{t+n} = \sum_{i=1}^{n} (r'_{t+i-1})^2/n, \tag{2}$$

where n (window size) is the number of points in the short window, and r' denotes the demeaning (i.e. the mean is removed) return. A shorter window would lead to a rough and spiky volatility that is hard to forecast, whereas a longer window causes new forecasts to be dependent on a greater number of previous values. However, an overly large window introduces too much smoothness, which may lead to underfitting volatility trends which occur over a finer time resolution.

In this paper n is set to be 10, which corresponds a time interval of 5 hours. The volatilities are normalised by removing the mean and scaling to unit variance:

$$V' = (1/\sigma)(V - \mu). \tag{3}$$

The normalised short-term realised volatility computed using Equation 3 for both positive and negative EUR/USD Forex returns is shown in Figure 2(c) and (d).

D. Gaussian processes

A Gaussian process is a collection of random variables such that every finite of collection of those random variables has a multivariate normal distribution. It is a popular technique in machine learning and is widely used in time series analysis [14]. Gaussian Process Regression (GPR) is a kernel-based nonparametric method that relies on the appropriate selection of a kernel and hyperparameters. For a given prior that defines the belief, GPR learns the posterior over the output space with observations [15].

For a regression of $\mathbf{y} = f(\mathbf{x})$ drawn from a multivariate Gaussian distribution, the dependent variable $\mathbf{y} = [y_1, y_2, ..., y_n]$ is evaluated at the set of locations $\mathbf{x} = [x_1, x_2, ..., x_n]$.

Gaussian processes are parameterised by a mean function $\boldsymbol{\mu}_*$ and a covariance function $\mathbf{C}_*$. The predictive distribution at $\mathbf{x}_*$, by using the standard results from multivariate Gaussian distribution, is given as:

$$\mathbf{p}(\mathbf{f}_*|\mathbf{x}, \mathbf{y}, \mathbf{x}_*) = \mathbf{N}(\mathbf{f}_*|\boldsymbol{\mu}_*, \mathbf{C}_*), \tag{4}$$

where:

$$\boldsymbol{\mu}_* = \mathbf{m}(\mathbf{x}_*) + \mathbf{K}(\mathbf{x}_*, \mathbf{x})\mathbf{K}(\mathbf{x}, \mathbf{x})^{-1}(\mathbf{y} - \mathbf{m}(\mathbf{x})), \tag{5}$$

$$\mathbf{C}_* = \mathbf{K}(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{K}(\mathbf{x}_*, \mathbf{x})\mathbf{K}(\mathbf{x}, \mathbf{x})^{-1}\mathbf{K}(\mathbf{x}, \mathbf{x}_*). \tag{6}$$

$\mathbf{m}(\mathbf{x}_*)$ and $\mathbf{K}(\mathbf{x}_*, \mathbf{x}_*)$ denote the prior mean function and the prior covariance function $\mathbf{K}$ respectively. $\mathbf{K}(\mathbf{x}, \mathbf{x})$ and $\mathbf{K}(\mathbf{x}_*, \mathbf{x})$ denote the covariance between input and itself, and the input and observations respectively. The conditional distribution of Equation 4 explains given a set of observed variables $\mathbf{y}$ at locations $\mathbf{x}$, what the belief of variable values at a set of testing locations $\mathbf{x}^*$ would be.

For noisy outputs with uncorrelated noise, a white noise hyperparameter is added to the diagonal of $\mathbf{K}$ by replacing $\mathbf{K}(\mathbf{x}, \mathbf{x})$ with $\mathbf{K}((\mathbf{x}, \mathbf{x}) + \sigma_n^2\mathbf{I})$.

The positive semi-definite covariance matrix $\mathbf{K}(\mathbf{x}, \mathbf{x})$ is a measure of how inter-related variables are within a vector. It describes the spatial covariance of a random variable process. The closer the data, the higher the covariance [4], [16].

a) Hyperparameter tuning: The kernel hyperparameters, namely the lengthscale, signal variance and noise variance, are interpretable and can be learned from data. The optimal set of hyperparameter values, $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3, ...]$, can be found by performing maximum a posteriori estimation (MAP). MAP finds the particular mean vector and covariance matrix that define the most likely multivariate normal distribution to result in the observed data.

The hyperparameter posterior, according to Bayes' rule, is:

$$p(\boldsymbol{\theta} \mid \mathbf{x}, \mathbf{y}, M_i) = \frac{p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}, M_i)p(\boldsymbol{\theta} \mid M_i)}{p(\mathbf{y} \mid \mathbf{x}, M_i)} \tag{7}$$

However, as full Bayesian inference is computationally intractable, a type II maximum likelihood approximation can be made, where the marginal likelihood $p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}, M_i)$ is maximized instead.

As a logarithmic function is a monotonically increasing function, when using MLE, the logarithmic marginal likelihood function for a zero-mean Gaussian process over the cost function has the form of:

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\mathrm{MLE}} &= \arg\max_{\boldsymbol{\theta}} \log p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) \\ &= -\frac{1}{2}\mathbf{y}^T\mathbf{K}^{-1}(\boldsymbol{\theta})\mathbf{y} - \frac{1}{2}\log|\mathbf{K}(\boldsymbol{\theta})| - \frac{N}{2}\log 2\pi.\end{aligned}$$

The first term shows how the model fits the observed target data, the second term penalizes the complexity of the model, and the third term is the normalization constant (see [14, Ch. 2] for the derivation).

This optimisation can be approached by various algorithms. For discussions on these it is helpful to refer to [14, Ch. 18, p.113]. In the work presented here, the MLE/MAP scheme is used as it offers a well-studied solution without the need for demanding computation and is found in all major Gaussian process software libraries. With `GPflow`, calling the optimiser in the absence of a prior performs type-II MLE, whereas specifying a prior performs MAP.

1) Multivariate non-coregionalized Gaussian processes: Multivariate non-coregionalised GPs have the same properties as the previously discussed univariate GP in Section III-D. Multivariate Gaussian processes are parameterized by a mean function $\boldsymbol{\mu}_*$ and a covariance function $\mathbf{C}_*$ as shown in Equation 5 and 6. However, the output data in this case will be a multidimensional vector instead of a 1-dimensional vector described in the previous section. The kernel function and hyperparameter tuning methods are unaltered, and the covariance matrix has the same parameters as that of a univariate GP. The supplementary information from the other parameters of the training dataset vector can improve the prediction of a target variable.

2) Multivariate co-regionalised Gaussian processes: For a multi-input and multi-output model, a co-regionalised Gaussian process model can to be applied to maximise predictive performance. The model exploits dependencies between the inputs, between the outputs, and between the inputs and outputs. Co-regionalisation explores the multivariate spatial cross-correlation through the set of co-regionalisation matrices.

The linear model of co-regionalisation states that all regionalised variables being studied are generated by the same set of physical processes acting additively at different spatial scales [17]. Consider a simple generative model where there is one latent function $\mathbf{u}(\mathbf{x})$ sampled from a Gaussian process of zero mean and a covariance of $\mathrm{cov}[\mathbf{u}(\mathbf{x}), \mathbf{u}(\mathbf{x}')] = \mathbf{K}(\mathbf{x}, \mathbf{x}')$.

The output functions $\mathbf{f}_1(\mathbf{x})$ and $\mathbf{f}_2(\mathbf{x})$ are defined as:

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix}\mathbf{f}_1(\mathbf{x}) \\ \mathbf{f}_2(\mathbf{x})\end{bmatrix}, \tag{8}$$

where $\mathbf{f}_1(\mathbf{x}) = \mathbf{a}_1^1\mathbf{u}^1(\mathbf{x})$, and $\mathbf{f}_2(\mathbf{x}) = \mathbf{a}_2^1\mathbf{u}^1(\mathbf{x})$. The covariance for $\mathbf{f}(\mathbf{x})$ is computed as:

$$\begin{aligned} \operatorname{cov}[\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}')] &= \mathbb{E}\{\mathbf{f}(\mathbf{x})[\mathbf{f}(\mathbf{x}')]^T\} \quad (9) \\ &- \mathbb{E}\{\mathbf{f}(\mathbf{x})\}[\mathbb{E}\{\mathbf{f}(\mathbf{x}')\}]^T. \quad (10) \end{aligned}$$

Collecting the terms, one can obtain $\operatorname{cov}[\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}')]$ as

$$\operatorname{cov}[\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}')] = \mathbf{B}\mathbf{K}(\mathbf{x}, \mathbf{x}') = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \mathbf{K}(\mathbf{x}, \mathbf{x}'), \quad (11)$$

where $\mathbf{B} = \mathbf{a}\mathbf{a}^T$, is the positive semi-definite co-regionalisation matrix. For the detailed derivation of the previous results, it is helpful to refer to [18].

The kernel derived from the linear model of co-regionalisation is a sum of the products of two covariance functions. One of these two functions models the dependency between outputs that is independent of the input vector $\mathbf{x}$ (the co-regionalisation matrix $\mathbf{B}$), and the second one models the input dependency that is independent of the covariance function $\mathbf{K}(\mathbf{x}, \mathbf{x}')$.

IV. METHOD

The implementation details for GP forecasting with all the models are given in Algorithm 1. This shows the rolling-window approach adopted in this paper. The univariate, multivariate and co-regionalised GPs each use a window size $N_w = 50$, with a prediction range of $N_p = 10$.

Algorithm 1 Rolling-window GP

```
function (μ, σ) = GPVOLATILITY(X, FX_Vol, N_w, N_p)
      ▷ (μ, σ, N_w, N_p) are the mean, variance, window size
  and prediction range respectively

    for i = 0, 1, ... length(FX_Vol) do
        X_train = X[i : i + N_w]
        Y_train = FX_Vol[i : i + N_w]
        X_total = FX_Vol[i : i + N_w + N_p]
        model.train(X_train, Y_train)
        if i%5 == 0: then                    ▷ every 5 windows
            θ ←model.optimize(θ)             ▷ h-param θ
        end if
        μ, σ = model.predict(X_total)
    end for

end function
```

$N_w = 50$ data in each window are inputted as the training data, and predictions for 10 steps into the future are obtained using the optimised hyperparameters. Increasing the size of the training data window gives more accurate predictions. However, the drawback of this is the increase in computational time. We therefore choose a window size of 50 to trade off computation time and error rate reasonably.

Instead of re-evaluating the hyperparameters at each forecasting step, we choose to update the hyperparameters based on the previous optimised values every 5 windows to improve the computation speed. Furthermore, the previous trained set of hyperparameters is used to initialize the training of the next hyperparameter set. The signal variance and noise variance are considered to lie within the range $[0.01, 10]$. The optimizer used here, `scipy.optimize.minimize`, uses a gradient descent approach. To overcome local points of inflexion, we find that re-initialising the hyperparameters with the previous best values empirically produced the best results.

TABLE I: Performance metric on positive EUR/CHF data using different kernels using a univariate GP rolling window approach for all the samples of the time-series.

	Matern 3/2	Matern 5/2	RBF
Median error	**0.285**	0.295	0.308

The volatility prediction is carried out using a Matérn $3/2$ kernel, a $5/2$ kernel, and a squared exponential. Table I shows that the Matérn $3/2$ outperforms other kernels according to the median error defined as defined in Equation 13 over the entire EUR/CHF time-series data. We therefore use the Matérn $3/2$ kernel over the smooth exponential kernel and the smoother twice differentiable Matérn $5/2$. This is because noisy financial data tends to be not infinitely differentiable, and also highly non-smooth, favouring the least smooth of these kernels.

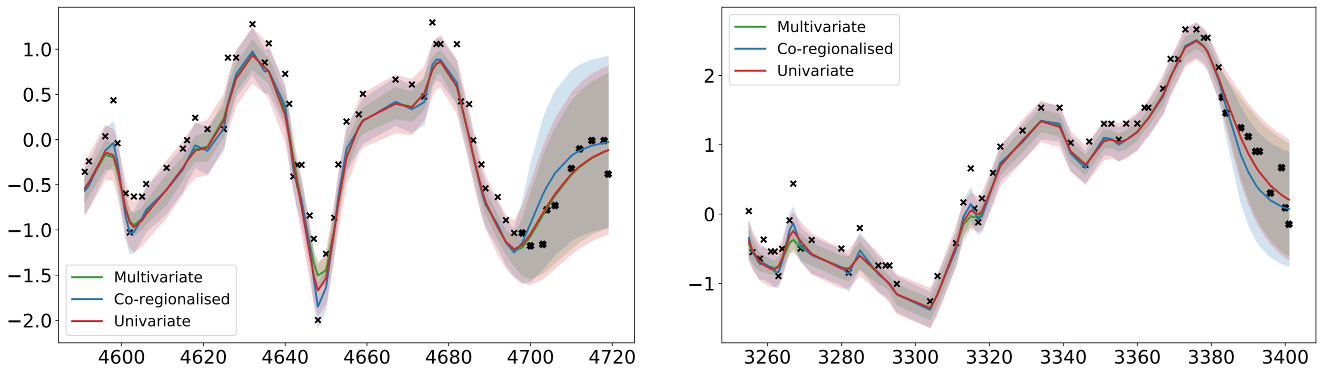
A. Multivariate GPs

The multi-dimensional input includes an additional dimension of the Eurodollar EUR/USD volatility data to the foreign exchange pairs we forecast, namely EUR/CHF. We then examine if the additional dimension of the input can give us more information. The joint posterior can provide a more robust and comprehensive result.

The multivariate GP learns the joint posterior distribution of each asset volatility using spatial covariance matrices specified by the 2D Matérn $3/2$ kernel. Both the active dimension and the Automatic Relevance Determination (ARD) options for kernel selection are deactivated as this allows the same kernel and hyperparameters set with a uniform weight to be applied in both dimensions.

B. Co-regionalised GPs

In a multi-output co-regionalised GP, the correlation between data streams are captured. In this scenario, the multiple training data sets are related and we make predictions on one data set based on a learnt correlation with a correlate data set. In the coregionalised GP model, the number of hyperparameters to optimize greatly increases due to the additional dimensions of the data . The base kernel in our case is Matérn $3/2$. The 'co-region' kernel indexes the outputs that act on the second data dimension. The GP model is constructed ant optimised in the same way as the univariate GP model.



(a) Univariate Positive GP EUR/CHF

(b) Univariate Negative GP EUR/CHF

Fig. 3: A comparison of univariate, multivariate, and co-regionalised GPs using positive and negative EUR/USD and EUR/CHF, illustrated for one window (of 5000) of the dataset. The training data are shown by thin markers and the test data are shown by thick markers. The aggregated error and predicitve uncertainty metrics over the corresponding one-step predictions are given in Table II.

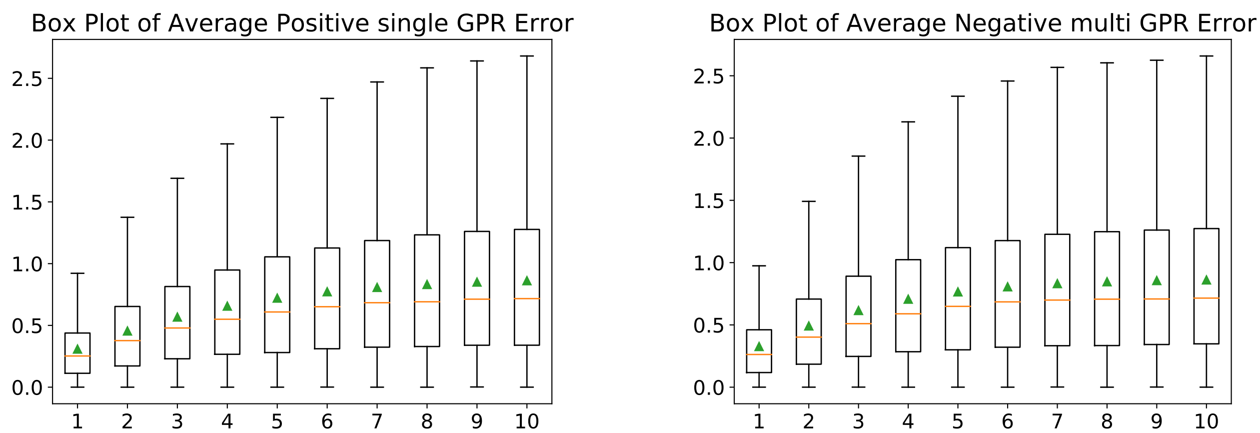


Fig. 4: 10-step ahead error of using a univariate (left) and multivariate Gaussian process model using the EUR/CHF data. A box plot depicts the errors through their quartiles. The diagrams give the first and third quartiles, the median (orange) and the mean (green triangle). We note that the error grows with the length of time that no new data point is observed, as expected.

V. Results

A. Univariate GPs

The performance of various forecasting techniques is evaluated by comparing the forecasted short-term realised volatility, as shown in Equation 2, to the true short-term realised volatility. The performance metrics used in this paper, the Mean Absolute Error (MAE) and Median Error, measure the forecasting performance, defined as:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\text{error}|, \tag{12}$$

$$\text{Median Error} = \text{median}(|\text{error}|), \tag{13}$$

where n is the number of forecasts, and the error at time t is defined as $\text{error}_t = \text{forecasted volatility}_t - \text{real volatility}_t$. Equation 12 gives a measurement of the average error and is affected by, whereas Equation 13 gives a more robust measure of the performance as the median is not affected by outliers.

The performance metrics are used to compare the forecasting power of various GP models for both one-step ahead and 10-step ahead predictions in the forthcoming scenarios.

Figure 3 shows a single window from the predictions performed with Algorithm 1 with all the GPs. All the models exhibit very similar behaviour over the regressed points, given by the thin markers, showing the ability of the GP to model financial volatility. In this example, the forecasted results (over the thick markers) using univariate and multivariate Gaussian processes are similar. However, the multivariate non-coregionalised GP volatility predictions have a smaller variance compared to the univariate and co-regionalised GP forecasts. This suggests the additional input does have the desired effect of reducing the uncertainty. We quantify this as the average standard deviation given by the shaded regions for each GP of Figure 3 effect in Table II to support this

TABLE II: Results of the rolling-window approach of Algorithm 1 performed on positive and negative volatilites of the EUR/CHF Forex rate. The predictions are made for one step ahead on a rolling window basis, with all error metricsresults reported out-of-sample.

	MAE	Median	Std. Deviation	Predictive Var.
Positive EUR/CHF volatility				
Univariate	**0.323**	**0.285**	**0.232**	0.648
Co-regionalised	0.538	0.437	0.433	0.679
Multivariate	0.330	0.286	0.244	**0.579**
Negative EUR/CHF volatility				
Univariate	**0.317**	**0.271**	**0.239**	0.683
Co-regionalised	0.532	0.432	0.444	0.751
Multivariate	0.327	0.277	0.258	**0.586**

statement.

Furthermore, we show the effect of the number of steps ahead that we predict blindly (in the absence of incoming data) in Figure 4. As one would expect, the variance grows proportionally to the length of time that no observation is received, due to the principled handling of model uncertainty in the GP framework.

The errors made with the rolling-window approach when forecasting a single step ahead for the entire data set are given in Table II. When measuring the performance using the median error, as well as the MAE, the univariate GP and multivariate non-coregionalised GP produce similar forecasting outcomes. We can conclude that the univariate GP and multivariate GP have indistinguishable volatility prediction power, however the multivariate GP is accompanied with predicting tighter variance bounds. Due to the increased complexity of the likelihood landscape associated with the co-regionalised GP, there is a greater likelihood of encountering local maxima, which affects the accuracy when making predictions. Consequently, the model fits the training data least favourably compared to the other GP models, with higher error and standard deviation, as well as higher predictive variance.

VI. Conclusion

This paper explores the application of Gaussian process regression in forecasting the volatility of foreign exchange (Forex) returns. We demonstrate the predictive ability of various Gaussian process models, namely the univariate, multivariate non-coregionalised, and multivariate co-regionalised GPs.

With a chosen kernel and window size, we infer hyperparameters using a type-II maximum likelihood method on the training data set that consists of different Forex pairs returns at 30-minute intervals. The n-step ahead short-term realised volatilities are forecasted using the trained hyperparameter sets, and the volatility predictions are tested with respect to the real short-term realised volatility.

The univariate and multivariate non-coregionalised GPs succesfully forecast EUR/CHF Forex data using a rolling-window one-step ahead prediction algorithm. Furthermore, the incorporation of the additional Forex time-series reduces consistently the uncertainties in the predictions made with the multivariate GP.

The performance of the co-regionalised GP model is limited by its large number of hyperparameters to optimise, which resulted in higher preditive error and variance consistently throughout all the tests made on the currency pairs. We recommend multivariate GPs if sufficient emphasis is placed on the posterior variance, such that the extra computational power can be justified. However, if the error in a least-squares or median sense is the of greatest importance, the computational budget may be better suited to more frequently re-tuning the hyperparameters of a univariate rolling-window GP.

References

[1] Paul Wilmott. *Paul Wilmott on Quantitative Finance-3 Volume Set*. John Wiley & Sons, 2006.

[2] Simon Benninga. *Financial modeling*. MIT press, 2014.

[3] Kenneth R French, G William Schwert, and Robert F Stambaugh. Expected stock returns and volatility. *Journal of financial Economics*, 19(1):3–29, 1987.

[4] Stephen Roberts, Michael Osborne, Mark Ebden, Steven Reece, Neale Gibson, and Suzanne Aigrain. Gaussian Processes for Time-series Modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984):20110550, 2013.

[5] Agathe Girard, Carl Edward Rasmussen, Joaquin Quinonero Candela, and Roderick Murray-Smith. Gaussian process priors with uncertain inputs application to multiple-step ahead time series forecasting. In *Advances in Neural Information Processing Systems*, pages 545–552, 2003.

[6] Sofiane Brahim-Belhouari and Amine Bermak. Gaussian process for nonstationary time series prediction. *Computational Statistics & Data Analysis*, 47(4):705–712, 2004.

[7] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

[8] Ser-Huang Poon and Clive WJ Granger. Forecasting volatility in financial markets: A review. *Journal of economic literature*, 41(2):478–539, 2003.

[9] Torben G Andersen, Tim Bollerslev, and Francis X Diebold. Parametric and nonparametric volatility measurement. In *Handbook of Financial Econometrics: Tools and Techniques*, pages 67–137. Elsevier, 2010.

[10] Nicolas Chapados and Christian Dorion. Gaussian Processes and Non-Parametric Volatility Forecasting. 2010.

[11] Yue Wu, José Miguel Hernández-Lobato, and Zoubin Ghahramani. Gaussian process volatility model. In *Advances in Neural Information Processing Systems*, pages 1044–1052, 2014.

[12] Syed Ali Asad Rizvi, Stephen J Roberts, Michael A Osborne, and Favour Nyikosa. A Novel Approach to Forecasting Financial Volatility with Gaussian Process Envelopes. *arXiv preprint arXiv:1705.00891*, 2017.

[13] Matthew Dixon and Stephane Crepey. Multivariate Gaussian Process Regression for Portfolio Risk Modeling: Application to CVA. *Department of Applied Mathematics, Illinois Institute of Technology*, 2018.

[14] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian Processes for Machine Learning*, volume 2. MIT Press Cambridge, MA, 2006.

[15] Zexun Chen. *Gaussian process regression methods and extensions for stock market prediction*. PhD thesis, Department of Mathematics, 2017.

[16] Marc Deisenroth, Yicheng Luo, and Mark van der Wilk. A Practical Guide to Gaussian Processes. `https://drafts.distill.pub/gp/`, 2019. Accessed: 2019-04-13.

[17] Michel Goulard and Marc Voltz. Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, 24(3):269–286, 1992.

[18] Mauricio A. Alvarez. Multiple-output Gaussian processes. `http://gpss.cc/gpss17/slides/multipleOutputGPs.pdf`, 2017. Accessed: 2019-05-03.