Gaussian Process Changepoint Detection in Finance

Tanay Trivedi TT1087@NYU.EDU
New York University

Yilun Kuang YK2516@NYU.EDU
New York University

Charlie Chen ZC2157@NYU.EDU
New York University

Abstract

Momentum strategies are widely used by portfolio managers as important parts of a successful portfolio management construction historically. These strategies struggled to adapt to rapidly evolving financial scenarios stemming from global crises, particularly during the COVID-19 market crash. Bayesian changepoint detection enabled by Gaussian Processes can help decide when old information must be thrown out and new trends should be identified. Building upon Wood et al. (2022), we apply the changepoint detection technique to a novel 7 liquid multilateral currency series dataset from the year 1990-2022 and show how changepoint detection can improve a classical momentum strategy. Furthermore, we explore empirical kernel selections and optimization procedures to identify inductive biases suitable for nonstationary financial time series. Lastly, we provide empirical analyses of the performance of different kernels via confusion matrix analysis, visualization of changepoints and induced functions.

1. Background

1.1. Momentum Strategies

Financial markets have predictable behavior for the majority of time, but can suddenly deviate from their known patterns for a short period and bankrupt traders[1] that rely upon steady patterns to make money. One such pattern that is used by traders is called the Momentum family of strategies, which invests on long-term assets that have been performing well in the past.

[1]This document will use traders to represent any active participant in financial markets if no other information is provided.

Proceedings of the 33^{rd} International Conference on Machine Learning, New York, NY, USA, 2016. JMLR: W&CP volume 48.

1.2. Users and Needs

Momentum trades proliferate the financial industry. On shorter time scales, it is used by high frequency traders on a microsecond basis, by intraday traders on a minutely basis, by hedge funds on a daily-weekly basis and by investment managers on a monthly basis. Observing this, it is obvious that patterns exist in the frequency domain of the financial markets, and the amount of money behind this investment thesis implies that many users will find the procedure we create useful.

1.3. Our Contribution

Based on the successful applications by (Wood et al., 2022), we will use the GP method for changepoint detection in financial data. More specifically, we intend to

1. Apply the GP method for changepoint detection from Wood et al. (2022) to a novel 9 liquid currency series dataset from the year 1990-2022.
2. Empirically compare the performance of different kernels (Spectral Mixture Kernel (Wilson & Adams, 2013), Matern family of kernels, compositions of Matern kernels) and identify the best inductive bias.
3. Develop an understanding via confusion matrix analysis and frequency spectrum visualizations of why certain kernels are successful at detecting changepoints in financial time series and why some that are well known in non-financial literature do not succeed in this setting

Our results from 1990-2020 show that with the proper kernel choices, it is possible to outperform baseline momentum models across the G9 currencies. Matern 1/2 kernel and their compositions with other kernels from the Matern family provides both wide (across currencies) and large (high information ratio) performance advantages. Spectral Mixture kernels with five mixture components performed on par with single Matern 3/2 and Matern 5/2 kernels across currencies.

In a holdout set from 2020-2022, where notably the COVID crash and other major economic events took place, the Composition of 1/2 and 5/2 continued to outperform, but other kernels that performed well in-sample from 1990-2020 underperformed.

2. Related Work

2.1. Bayesian Online Changepoint Detection

Changepoint detection is the task to find abrupt changes in parameters like mean, variance, and variable correlation retrospectively. Equivalently, the task can be phrased as partitioning the data into different segments, within which data share same parameters. Specifically, consider a time sequence $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_T\}$. It can be partitioned into disjoint contiguous segments $\{\rho_1, \ldots, \rho_N\}$.

Though there have been several frequentist attempt to solve this problem, Bayesian methods are suitable for this problem because it automatically finds a balance between the number of changepoints (model complexity) and model fit (Xuan & Murphy, 2007).

Mainstream changepoint models assume a prior over the length of each segments (Adams & MacKay (2007), Xuan & Murphy (2007), Fearnhead & Liu (2007)). In Adams & MacKay (2007), a hazard function $H(\tau)$ is defined and the conditional run length for the segment is

$$P\left(r_t \mid r_{t-1}\right)=\begin{cases}H\left(r_{t-1}+1\right) & \text{if } r_t=0 \\ 1-H\left(r_{t-1}+1\right) & \text{if } r_t=r_{t-1}+1 \\ 0 & \text{otherwise}\end{cases} \tag{1}$$

A common choice for the prior on the segment length is geometric distribution. The corresponding hazard function is $H(\tau) = 1/\lambda$. With this model assumption, dynamic programming can be used to efficiently simulate the posterior distribution for changepoints exactly. Adams & MacKay (2007) presents a method to find changepoints while predict next data.

$$P\left(x_{t+1} \mid \boldsymbol{x}_{1:t}\right)=\sum_{r_t} P\left(\boldsymbol{x}_{t+1} \mid \boldsymbol{x}_t^{(r)}, r_t\right) P\left(r_t \mid \boldsymbol{x}_{1:t}\right) \tag{2}$$

Fearnhead & Liu (2007) simulates the set of changepoints in the reverse direction, starting from the last changepoints and then finding the one before that

$$P\left(r_n \mid \mathbf{x}_{1:n}\right) \tag{3}$$

Xuan & Murphy (2007) generalizes the method of Fearnhead & Liu (2007) and considers the multivariate case by utilizing the Gaussian graphical models (GMMs) to model the correlation between different variables. Saatçi et al. (2010) argues that in Adams & MacKay (2007), the data points in each segment are i.i.d, which doesn't well reflect the reality, so they use Gaussian Process to model the function in each segment.

For the methods described above, a prior on the run length is required as a hyperparameter, which largely affects the performance. Also, in the financial time series, in spite of the existence of cyclic growth, the volatility of the market is usually unstable and unpredictable, especially under some unexpected events like 2008 Financial Crisis and outbreak of COVID19. So, it may not be reasonable to assume a typical value for run length.

In the changepoint modeling in Wood et al. (2022), they use the method proposed in Lloyd et al. (2014), which is also the method we are using because it doesn't have an assumption for the run length and is flexible enough to capture the changepoints with well-designed kernel.

2.2. Changepoint Detection in Finance

Besides human intuition, the only known technology studied in financial literature is the use of Hidden Markov Models, which is a probabilistic state transition model that identifies states of data which maximize the difference between studied variables. Once these states are identified, the algorithm can compute the most likely state that any observation belongs to. Formally, given a discrete-time stochastic process (X_n, Y_n) where Y_n is the observation and X_n is some hidden latent variable ($X_n \in \{0, 1, 2, \ldots\}$), $P(X_n = 1)$ will tell whether the observation Y_n belongs to state 1. In the financial application, percent returns of some asset or strategy are the observations Y_n and the latent states are market regimes. The technique is somewhat successful in a backwards looking context, but when tested out of sample were not able to avoid the market fallout from the COVID-19 lockdown. This is where online change point detection offered by Gaussian Processes can help. Past trained models do not retain the continuous training characteristics of online methods.

3. Methodology

3.1. Gaussian Process for Changepoint Detection

Consider a sequence of daily returns $\hat{\boldsymbol{r}} := \{\hat{r}_{T-\ell}, ..., \hat{r}_T\}$ where $\hat{r}_t = (r_t - E[r_t])/\sqrt{\text{Var}(r_t)}$ and ℓ is the length of a fixed look-back window. Following Wood et al. (2022), we consider the generative model

$$\hat{r}_t = f(t) + \epsilon_t, f \sim \mathcal{GP}(0, k_\xi), \epsilon \sim \mathcal{N}(0, \sigma_n^2).$$

The covariance function k_ξ is typically chosen to be the Matérn 3/2 kernel parametrized by ξ as standard practices. Matérn 3/2 kernel is known to induce non-smooth function draws (Wood et al., 2022).

To identify changepoint, Wood et al. (2022) construct the following kernel

$$k_{\xi_C}(x, x') = k_{\xi_1}(x, x')\sigma(x, x') + k_{\xi_2}(x, x')\bar{\sigma}(x, x') \quad (4)$$

where $\sigma(x) := 1/(1 + e^{-s(x-c)}), \sigma(x, x') = \sigma(x)\sigma(x')$, $\bar{\sigma}(x, x') = (1-\sigma(x))(1-\sigma(x'))$, and the potential changepoint c for $c \in (t - \ell, t)$ (Wood et al., 2022). Intuitively, σ and $\bar{\sigma}$ provide a smooth transition between two different kernels. In the case where k_{ξ_i} is from the Matern family of kernels, the changepoint kernel k_{ξ_C} allows us to identify a break point between two stationary regimes and thus capture non-stationarity in the data.

To perform changepoint detection, Wood et al. (2022) optimize the log marginal likelihood with respect to $\xi_C = \{\xi_1, \xi_2, c, s, \sigma_n\}$ to obtain the optimal change point $\hat{c}$.

3.2. Multi-Currency Seeding + Fine Tuning

Given fixed kernels, the hyperparameters are optimized by the GPFlow, which uses `scipy.optimize.minimize` under the hood. The data is first fitted with data without changepoints and then with the changepoint kernel described in formula 4. The score for the chagepoints is then (Wood et al., 2022)

$$\nu_t^{(i)} = 1 - \frac{1}{1 + \mathrm{e}^{-\left(\mathrm{nlmn}_{\xi_C} - \mathrm{nlmn}_{\xi_M}\right)}} \quad (5)$$

where nlmn_{ξ_C} is the negative log marginal likelihood for the naive kernel and nlmn_{ξ_M} is the counterpart for the changepoint kernel.

Notice that the log marginal likelihood with respect to the changepoint location c is not convex. So the initialization point for hyper parameters may influence the performance by a large margin. To obtain a schematically meaningful initialization, we try to first optimize kernel hyperparameters with all currencies as vector because the model may capture the structural relationship between multiple currencies and give us more reliable changepoints. Based on this initialization, we further finetune the hyperparameters on each individual currency to find the specific changepoints for the respective currency. We refer to this as *Multi-Currency Seeding + Fine Tuning*.

3.3. CPD Resolution Methods

Running the GP models returns a set of changepoint locations and their score. However, since changepoints are located by different windows of date independently, the returned changepoints are cluttered on some certain dates. To avoid overcounting the same changepoints many times and not missing them, a resolution strategy is adopted. There is a two-step filtering procedure. First, changepoints lower than some threshold t will be filtered out. Second, within some window of days τ, only the day with the largest score will be selected. $\{t, \tau\}$ can be adjusted to obtain a resonable number of changepoints.

In our experimentation, we used Section 4.2 with settings $t = 5, \tau = 0.98$. This results in about 30 changepoints a year on average.

3.4. Evaluation

Changepoint detection is unique in the world of time series machine learning, because its labels are not easily identifiable. Usually, the output must be hand evaluated to check its accuracy. In this situation, a business logic can be used to grade the changepoint detection. A trading portfolio can be directly built on top of the identifcation of a new changepoint. A simplistic algorithm for a single currency is given below:

1. Start with a position $P = sgn(M)$ where M is the average return of the currency over a 90d window horizon and an empty list $P_{list} = [], R_{curr} = []$
2. Loop through the dates, recording new changepoints generated by the GP model at each step
   (a) At each changepoint, flip the sign of P
   (b) Append P to P_{list}; append the revenue to R_{curr}
3. $R = P_{list} \cdot R_{curr}$ where R represent the revenue operated by the strategy solely based on the detected changepoint information.

The output portfolio return R can be compared to the benchmark, which is to hold a cross-sectional momentum based portfolio for that currency.

On performance metrics, the primary one is Information Ratio $IR = \frac{\mu_r}{\sigma_r}$, where μ_r is the average return of the strategy and σ_r is the standard deviation of returns of the strategy.

4. Experiments and Discussion

4.1. Dataset

Currencies are traded in pairs. Usually, a portfolio manager wanting to buy Euros will use some other currency they already hold to buy Euros, but their returns will always be reported in one base currency (commonly USD). So in this project, we will be using USD pairs against 8 other currencies (G10 without NOK as institutional investors do not commonly trade energy based currencies).

A standard practice borrowed from the realms of algorithmic trading and machine learning is to separate some holdout data to test hyperparameter and model choices on. In our case, we have reserved 2020-2022 as our holdout data. Results below are separated as in-sample and holdout set, as we are restricting that for testing and reporting on our final paper. Until then, the holdout data was unseen and we did not iterate or tune manually or automatically on it.

4.2. Kernel Selection

For numerical experiments, we use the change point kernel module in GPflow on our currency data with a variety of stationary covariance functions (Matthews et al., 2017).

More specifically, Matern 1/2, 3/2, 5/2 kernels are used due to their compatibility with financial time series (Liu et al., 2020). Additionally, we adapted implementations of the Spectral Mixture (SM) kernel (Gadicherla, 2018) since SM can ideally represent all stationary covariance function as it's derived from the Bochner Theorem (Wilson & Adams, 2013). As the Matern 1/2 kernel is the most volatile, we expect it to complement the classical Matern 3/2 kernel in market data (Rizvi, 2018; Wood et al., 2022) and thus also include a sum of Matern 1/2 and Matern 3/2 kernel. We also include a sum of Matern 1/2 and Matern 5/2 kernel.

For initialization, Matern 1/2, Matern 3/2, and Matern 5/2 have a default initialization of 1.0 for both length scale and variance. The SM kernel mixture weights are chosen to be the standard deviations of the target values divided by the total number of mixtures, which is fixed to be either 5 or 10. The mixture mean is drawn from Unif$(0, 0.5/\min|x-x'|), \forall x, x'$ in our datasets. The inverse of the length scale is given by the $\mathcal{N}(0, \max(|x-x'|)), \forall x, x'$.

5. Results

5.1. In-Sample Results

5.1.1. Discussion of Tables

Using the resolution methods discussed in Section 3.2 and the evaluation methodology given in Section 3.3, we evaluated the choices given in Section 4.2. These choices were running across the 21, 63, 126, 252 day lookback windows. Grouping the kernels as the basic Matern kernels (1/2, 3/2, 5/2), the Composition of Matern Kernels (1/2 + 3/2 and 3/2 + 5/2), the Spectral Mixture (5 and 10 kernels), and Pre-trained kernels, results are quoted in Table 1 above quoting the number of currencies out of 7 that outperformed the benchmark across lookbacks and the maximum IR across currencies and lookbacks.

Its clear that the composition of Matern 1/2 and 5/2 kernels outperforms across currencies and at the maximum Kernel IR. Its also interesting to note that amongst basic kernels, 1/2 and 5/2 outperform the 3/2. The composition of them brings out the best in each, combining their results to outperform on almost all currencies.

The original publication on financial Gaussian Process detection (Wood et al., 2022) indicated that the Matern 3/2 Kernel was the best choice for their experiments. We put down our differences in results to theirs to our data differences. Wood et al. (2022) was operating on financial futures contracts covering everything from major stock market indices to agricultural commodities, whereas our focus is much tighter on just the G9 currencies. As our universe is smaller and more homogenous, it is not surprising that the kernel choice that we found best is not the most general middle path that the cited paper chose to pursue. For us, the Matern 1/2 and 5/2 out perform the 3/2, and the composition of both outperform the single kernels, while the 3/2 seems to be too general to capture either short term volatility (1/2) or long term trends (5/2).

Interestingly, Spectral Mixture (SM) kernel outperforms single Matern 3/2 and Matern 5/2 kernels by 0.08 in IR, contradicting conventional wisdom of the usages of these two kernels in financial markets (Rizvi, 2018; Wood et al., 2022). We suspect this is because the SM kernel forms a basis for all stationary covariance functions (Wilson & Adams, 2013), and data-driven learning of the hyperparameters of the SM kernel induces a kernel with volatility roughly the same as that of Matern 1/2 compared to Matern 3/2 and Matern 5/2 kernels. In future study, we intend to investigate this hypothesis by comparing to function draws from the learned Matern 1/2 kernel and the learned SM kernel.

5.1.2. Out-of-Sample Results

In Table 2, the results for the 2020-2022 holdout data are reported.

The rough patterns in-sample extend to out-of-sample holdout data as well. The composition of Matern 1/2 and 5/2 outperform on the majority of currencies.

The performance difference is not as stark in terms of currencies outperforming in the holdout set. This period is

Family of Kernel	Kernel Settings	Currencies Outperforming	Max Kernel IR	Max Baseline IR	Number of CPDs
Basic Kernel	1/2	4	0.34	0.17	381
	3/2	3	0.26	0.17	402
	5/2	4	0.26	0.17	453
Composition of Kernels	1/2 + 3/2	2	0.32	0.17	420
	1/2+5/2	7	0.36	0.17	427
Spectral Mixture	5 Kernels	4	0.17	0.17	2232
	10 Kernels	0	0	0.17	2231
Prior Training on Currencies	1/2	2	0.33	0.17	412
	3/2	4	0.22	0.17	421

Table 1. Results of G9 Currency Experiment In Sample 1990-2020

Family of Kernel	Kernel Settings	Currencies Outperforming	Max Kernel IR	Max Baseline IR	Number of CPDs
Basic Kernel	1/2	2	0.75	1.03	30
	3/2	3	0.85	1.03	31
	5/2	2	0.85	1.03	49
Composition of Kernels	1/2 + 3/2	1	1.24	1.03	31
	1/2+5/2	4	1.24	1.03	47
Spectral Mixture	5 Kernels	1	0.82	1.03	216
Prior Training on Currencies	1/2	3	1.05	1.03	40
	3/2	3	1.05	1.03	37

Table 2. Results of G9 Currency Experiment Out of Sample 2020-2022

particularly difficult to work with, as the COVID crisis and ensuing economic gyrations are difficult for models to deal with. Its not surprising therefore that the model only outperforms on 4, rather than 7 currencies as before. The IRs are higher in this table simply because this period is smaller and more volatile than most of the earlier in-sample 1990-2020 period, so the large swings produce bigger opportunities to make money.

5.2. Visualization of Changepoint Kernel

One fair question to ask is that if the changepoint kernel is actually useful to capture the change in probability distribution and if capturing this changepoint information is useful to the predictive distribution and modeling of the data. To show this, we plot one of the changepoint kernels recovered from the saved parameters during training by drawing several functions from the kernel in Figure 1. In the figure, we can see that the drawn functions has a clear change in variability before and after the origin, demonstrating that the changepoint kernel is active in separating the time period into two different regimes.

Next, we want to investigate the effectiveness of the changepoint kernel versus the naive kernel. In Figure 2, given a look-back-window, we show the predictive distribution by two different kernels. On the right hand side, we see that the Matern kernel doesn't capture the change in stationarity, resulting in a poor estimate for the returns. On the other hand, the changepoint kernel on the left captures the change in stationary regimes. From visual figures, the changepoint kernel clearly models the data better than the naive kernel. Thus, we prove that the changepoint kernel can capture the changepoint information that separates two stationary regime and with this information, the Gaussian Process has more modeling power for financial series data.

5.3. Visualization of changepoints

For an example of how to visualize the detection of changepoints on the currency data, observe the results on the Swiss Franc through the Great Financial Crisis. Figure 3 is a plot of the currency rate with vertical lines for the changepoints.

The largest market regime change is through 2008H1, where the Swiss Franc depreciated 10% against the US Dollar before reversing back the losses. Its clear that the changepoint detection procedure catches this loss and also

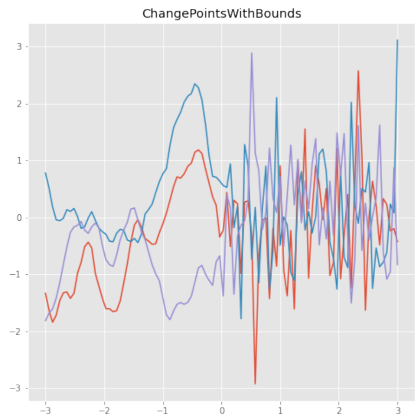

Figure 1. Changepoint Kernel Function Sampling
The parameter to recover this kernel is retrieved from the training period. For illustrative purpose, the changepoint location is changed to be at the origin. Different sampled parameters both recover kernel that demonstrate different distribution before and after the changepoint, so only one example is shown here.

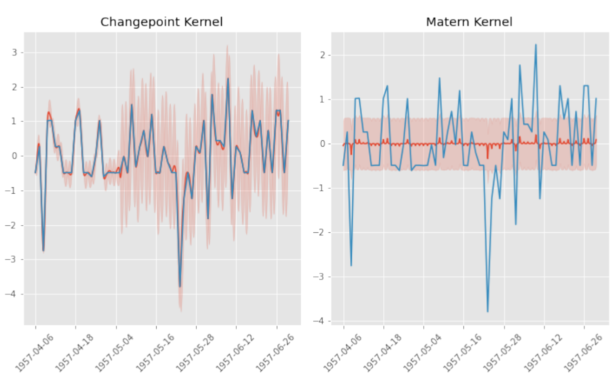

Figure 2. Predictive Distribution by Different Kernels
The look-back-window is 63 days and the base kernel for the changepoint kernel is Matern32. The blue line is the line of data. The red line is the predictive mean. The red shadow is the region up to two standard deviations.

identifies part of the reversion as well, since the vertical lines bound the region and don't incorrectly find superfluous changepoints within the region.

5.4. Confusion Matrix and Analysis

Table 3 is the confusion matrix for the best model in-sample. The confusion matrix is constructed by taking the sign of the position in between the changepoint switch as the prediction and the sign of the average return between the changepoint switch as the label.

The accuracy in terms of sign is not very high, it hovers a little above 50% for most of the currencies meaning that its not a very good predictor of the sign of the average return. Based on the comparison of this and the information ratio results, it can be seen that the models are capturing large swings and making money on them, rather than attempting to find every possible stationarity change.

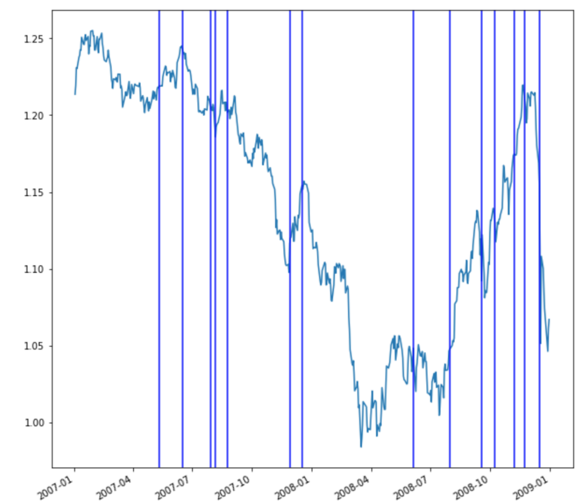

Figure 3. Change points on Swiss Franc Through GFC

A more complete confusion matrix analysis would also have some labels for directly identifying changepoints. An offline stationarity shift finding algorithm could be used on the entire series at once, and compared to the online algorithm presented in this paper. We leave this as future work.

5.5. Function Space Analysis of Spectral Mixture vs. Matern Kernel during COVID Market Crash

In Table 1 and Table 2, we notice that the Spectral Mixture kernel differs from the Matern family of kernels in the number of changepoints, max kernel information ratio, and the number of outperforming currencies compared to the baseline method. Given that the Spectral Mixture kernel ideally is able to represent the stationary Matern family of kernels, it's worth investigating its failure mode.

fig. 4 compares function draws from the Spectral Mixture kernel and the best performing sum of Matern 12 and Matern 52 kernel. We can see that the induced functions of the Spectral Mixture kernel have some phase match but amplitude mismatch compared to the function draws from the sum of Matern kernels. Thus the learned Spectral Mixture kernel may have overestimated the periodicity and volatility of currency time series.

In addition, we plot the frequency spectrum of the induced functions of the learned Matern 12, Matern 52, and Matern 12 + 52 in fig. 5. Interestingly, the spectrum of Matern 12

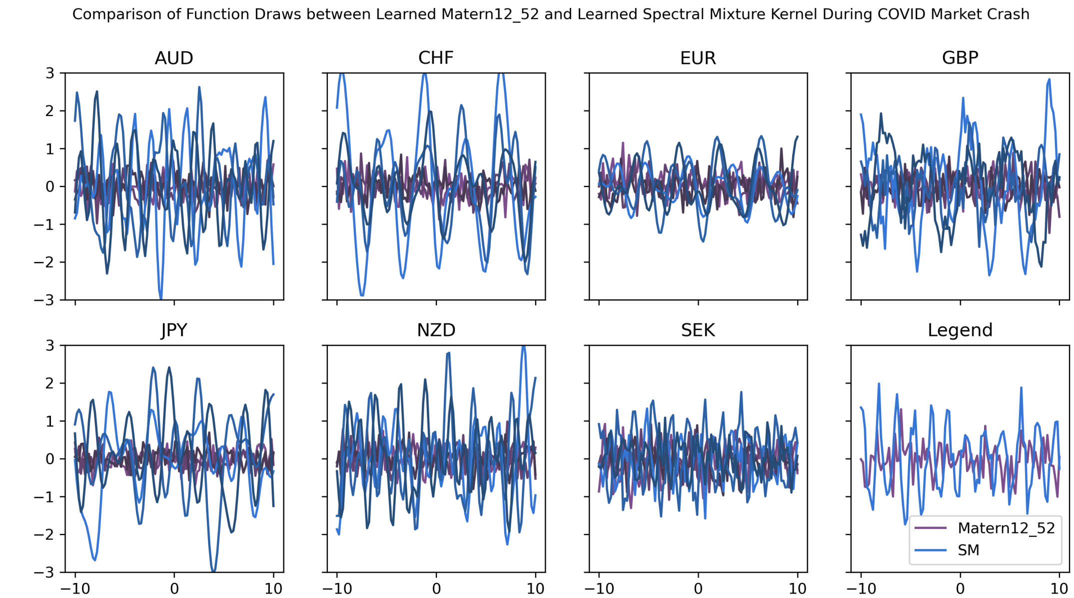

Figure 4. The function draws from the learned Matern12 + Matern52 sum kernel and the learned Spectral Mixture kernel during the 2020-2022 COVID Market Crash are compared against over seven currencies. The last figure is included for labeling purposes.

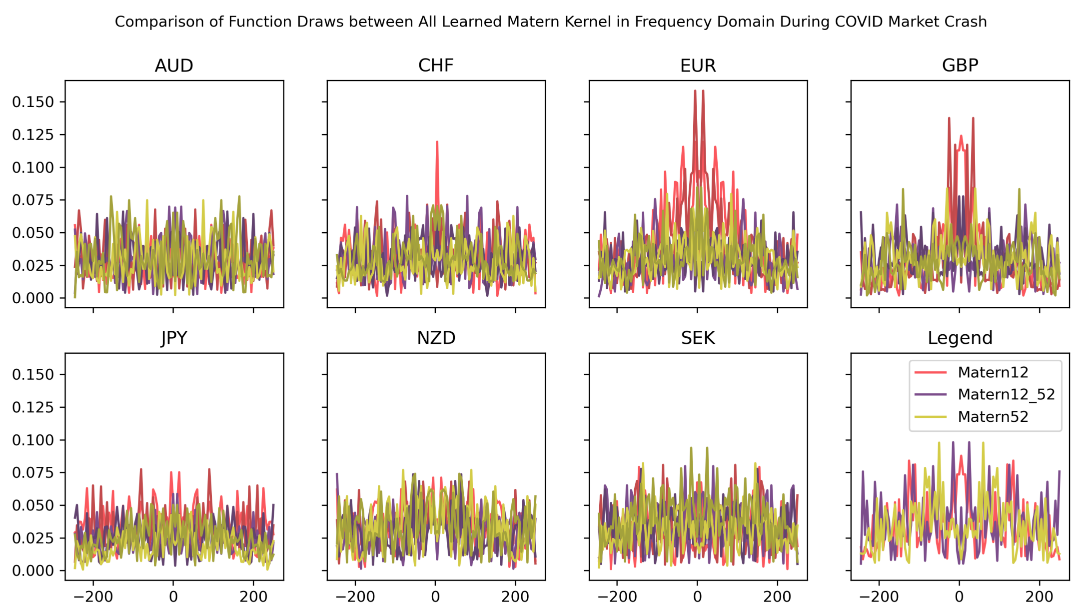

Figure 5. The frequency spectrum of function draws from the learned Matern12, Matern52, and Matern12 + Matern52 during the 2020-2022 COVID Market Crash is compared against over seven currencies. The last figure is included for labeling purposes.

Currency	Real Market Direction	Predicted Upmarket	Predicted Downmarket
AUD	Up	42%	7%
	Down	40%	9%
EUR	Up	23%	17%
	Down	23%	17%
GBP	Up	38%	13%
	Down	24%	25%
SEK	Up	44%	6%
	Down	45%	5%
JPY	Up	2%	48%
	Down	2%	48%
CHF	Up	8%	41%
	Down	8%	42%
NZD	Up	22%	29%
	Down	21%	29%

Table 3. Confusion Matrix for the best model across currencies

concentrates in the low frequency regime. This is potentially because the optimal lookback window for the Matern 12 kernel is 252 days. Compared to the Matern52 and Matern 12 + Matern 52 with 21 days of lookback window, the low frequency 252 days could be relatively high frequency if we compress 252 to 21.

For the same lookback window of 21 days, the spectrum Matern 52 and Matern 12 + Matern 52 is qualitatively similar except that the sum of Matern 12 and Matern 52 tends to have more density in the high-frequency regime and also possesses comparable density in the low-frequency regime. Given that the sum of Matern 12 and Matern 52 performs the best in both the in-sample 1990-2020 and out-of-sample 2020-2022 period, We suspect that the combination of these two captures both breaks in stationarity signified by the high frequency power and the momentum trend manifested by the low-frequency regime.

6. Conclusion

In this paper, we investigate the effectiveness of the strategy induced by the changepoint detection model on a multi-currency datasets. We use Gaussian Process with changepiont kernel to find the changepoints and revert the position whenever there is a changepoint. We show that it beats the traditional momentum methods, which is commonly used by traders today. Moreover, we empirically experiment the Matern family and Spectral Mixture kernels and find that Matern12 + Matern52 is the best combination, beating the common choice of Matern32.

By numerous visualizations, we demonstrate that the changepoint kernel effectively captures changepoint information and enhances the data modeling process, leading to a strategy that achieves positive returns and outperforms traditional momentum methods. We also found that the Matern12 + Matern52 kernel yielded the best and consistent results across both the training and testing sets. We notice that in all the position reversion induced by the changepoints, the accuracy is relatively low possibly because there is a lot of noise in the detected changepoints, which include false positives. Since our current strategy based on changepoint locations is relatively simplistic, there is a lot of room to improve and refine our strategies that leverage changepoints in a more effective way. We leave this as future works.

References

Adams, Ryan Prescott and MacKay, David JC. Bayesian online changepoint detection. *arXiv preprint arXiv:0710.3742*, 2007.

Fearnhead, Paul and Liu, Zhen. On-line inference for multiple changepoint problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4): 589–605, 2007.

Gadicherla, Srikanth. Spectral mixture kernel in gpflow. 2018. URL https://github.com/imsrgadich/gprsm.

Liu, Bingqing, Kiskin, Ivan, and Roberts, Stephen. An overview of gaussian process regression for volatility forecasting. In *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pp. 681–686, 2020. doi: 10.1109/ICAIIC48513.2020.9065045.

Lloyd, James, Duvenaud, David, Grosse, Roger, Tenenbaum, Joshua, and Ghahramani, Zoubin. Automatic construction and natural-language description of nonparametric regression models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28, 2014.

Matthews, Alexander G. de G., van der Wilk, Mark, Nickson, Tom, Fujii, Keisuke., Boukouvalas, Alexis, León-Villagrá, Pablo, Ghahramani, Zoubin, and Hensman, James. GPflow: A Gaussian process library using TensorFlow. *Journal of Machine Learning Research*, 18(40):1–6, apr 2017. URL http://jmlr.org/papers/v18/16-537.html.

Rizvi, SAA. *Analysis of financial time series using nonparametric Bayesian techniques*. PhD thesis, University of Oxford, 2018.

Saatçi, Yunus, Turner, Ryan D, and Rasmussen, Carl Edward. Gaussian process change point models. In *ICML*, 2010.

Wilson, Andrew and Adams, Ryan. Gaussian process kernels for pattern discovery and extrapolation. In *International conference on machine learning*, pp. 1067–1075. PMLR, 2013.

Wood, Kieran, Roberts, Stephen, and Zohren, Stefan. Slow momentum with fast reversion: A trading strategy using deep learning and changepoint detection. *The Journal of Financial Data Science*, 4(1):111–129, 2022.

Xuan, Xiang and Murphy, Kevin. Modeling changing dependency structure in multivariate time series. In *Proceedings of the 24th international conference on Machine learning*, pp. 1055–1062, 2007.

Appendix

A. Difference between Midterm and Final

Specifically, we plotted the functions drawn from the trained kernels, demonstrating that the regime before and after the changepoints are different. We drew the predictive distribution for the time window, contrasting the performance of changepoint kernels and naive ones. And the changepoint kernel significantly beat the naive Matern family.

We also calculate the win-loss ratio of the changepoints. This ratio essentially measures the probability of earning money whenever we flip the position based on the changepoints.

B. Additional Details on Kernel Comparisons

This section provides additional figures on function draws comparison between the Spectral Mixture Kernel and the Matern 12, Matern 52 kernels. We can see that the Spectral Mixture Kernel consistently induces functions of larger amplitude.

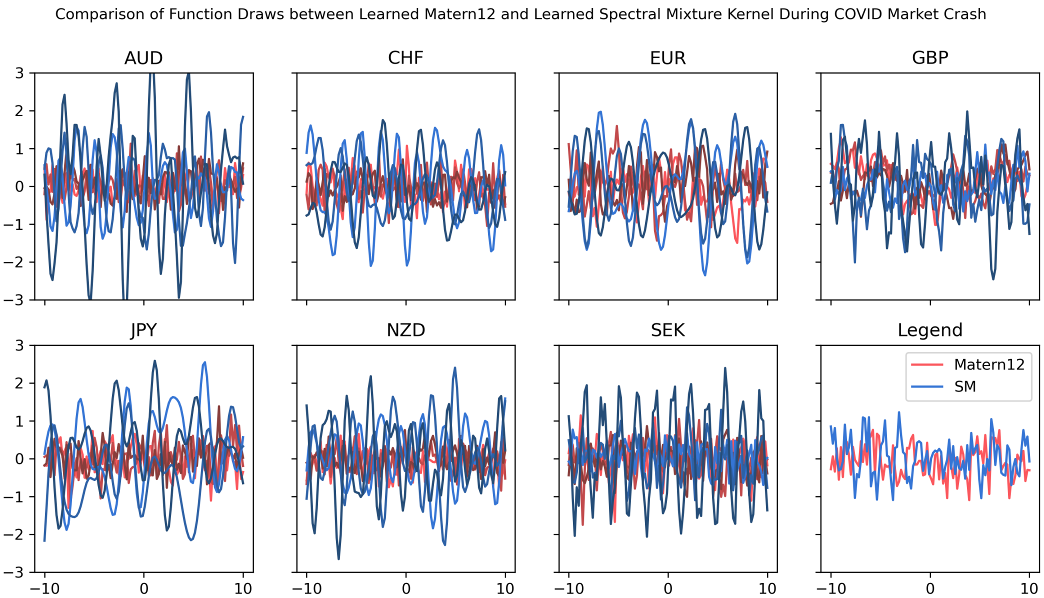

Figure 6. The function draws from the learned Matern12 kernel and the learned Spectral Mixture kernel during the 2020-2022 COVID Market Crash are compared against over seven currencies. The last figure is included for labeling purposes

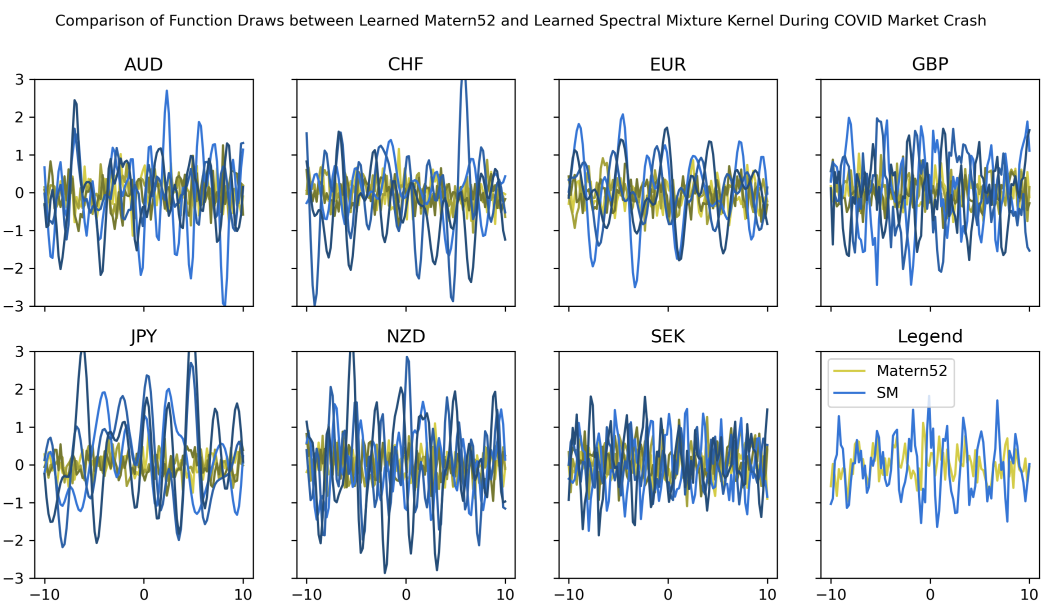

Figure 7. The function draws from the learned Matern52 kernel and the learned Spectral Mixture kernel during the 2020-2022 COVID Market Crash are compared against over seven currencies. The last figure is included for labeling purposes