**Automatic change-point detection using piecewise support vector quantile regression**

Jing-Rung Yu1, Sheng-Yu Tsai1, Chun-Yen Tsai1, Donald Lien2

[jennifer@ncnu.edu.tw](mailto:jennifer@ncnu.edu.tw), justintsai77@gmail.com, steven1115599@gmail.com, Don.Lien@utsa.edu

Department of Information Management

National Chi Nan University

Nantou, Taiwan1

Department of Economics

University of Texas at San Antonio, USA2

Abstract

Pattern change detection (PCD) in time series data is essential for identifying structural shifts, enabling accurate forecasting and informed decision-making across domains such as finance, healthcare, IoT, and manufacturing. This study proposes a piecewise support vector quantile regression (PSVQR) model to achieve automatic change-point detection and accurate interval prediction. The PSVQR model combines four key components: it employs the quantile loss function to estimate conditional quantiles, generating prediction intervals that account for non-constant variance and data heterogeneity; it incorporates a piecewise structure formulated through multi-objective programming to capture structural changes by segmenting data into distinct regions; it enables automatic detection of multiple change-points by treating their number as an optimization objective; and it utilizes feature selection to refine input variable selection and improve model parsimony. Overall, the PSVQR framework offers a flexible, adaptive, and computationally efficient solution that outperforms conventional methods in handling complex datasets with structural shifts.

Keywords: Change-point detection, SVR, Quantile loss, Piecewise, Feature selection

1. **Introduction**

Since Page's seminal work in 1955, testing for parameter changes has been a critical issue across fields like economics, engineering, and medicine, leading to extensive research (Csörgö & Horváth, 1997). Nowadays, pattern change detection (PCD) is essential across fields such as finance, healthcare, IoT, and manufacturing, where identifying shifts in sequential data patterns is crucial. In healthcare, PCD helps identify emerging health issues, such as sleep disorders, through altered sleep patterns or changes in physical activity levels. Financial fraud detection benefits from identifying unusual transaction behaviors, such as changes in credit card usage frequency or location (Pourhabibi et al., 2020). Sequential data plays a central role in IoT devices like sensors, which generate continuous data to uncover patterns and anomalies (Sgueglia et al., 2022). In manufacturing, detecting issues like equipment failures by analyzing shifts in machine vibration patterns helps enhance system performance and ensure system security (Guo et al., 2019; Gupta et al., 2023). Properly managing PCD prevents missed intervention opportunities, false alarms, and misinterpretations.

Gupta et al. (2024) comprehensively analyzed change-point dynamics detection in time series data. PCD methods are generally categorized into parametric approaches, non-parametric approaches, and hybrid approaches. No single approach is universally effective for detecting pattern changes in sequential data. The performance of any method depends on the dataset’s specific characteristics. However, hybrid methods often outperform others by combining accuracy and versatility, enabling efficient identification of complex changes across diverse datasets. Despite higher computational cost, these methods adapt well to varying data features, making them suitable for a wide range of applications. Selecting an appropriate PCD algorithm should align with the unique requirements of the task and the nature of the data.

Gupta et al. (2024) concluded that key challenges in PCD include detecting multiple change-points, handling high-dimensional or missing data, and processing large datasets efficiently. Future work should focus on improving computational efficiency, integrating PCD with concept drift, and developing hybrid methods to enhance accuracy and reduce false detections. They also noted that quantifying uncertainty, such as confidence intervals, is more challenging in non-parametric settings than in parametric ones.

Time series forecasting problems are often reformulated as regression-based predictive modeling tasks (Jadon et al., 2024). However, the inherent complexity of time series data, such as seasonality and autocorrelation, limits the effectiveness of simple linear methods less effective. To tackle the challenges of identifying multiple change-points and managing uncertainty, this study aims to develop a piecewise support vector quantile regression (PSVQR) method. This approach will enable automatic change-point detection while providing interval predictions based on quantiles. The model simultaneously detects multiple change-points (Yu et al., 2001) using a multi-objective programming approach. Moreover, the *l*1-norm will be employed to improve resilience to outliers, while feature selection will be performed by optimizing the number of input variables and change-points. A kernel-free model will be incorporated to enhance explainability. Change-point neighborhood evaluation (CPNE) will be used to compare the neighborhood of the detected change-points (Svoboda et al., 2024) generated by the proposed method and benchmark models.

**1.1 Loss functions and why quantile loss function**

In recent years, the loss function has become a central focus in machine learning due to its significant theoretical and practical value (Wang et al., 2020). Selecting an appropriate loss function, also referred to as an objective function, is crucial for large-scale time series forecasting because it drives the algorithm’s learning process (Jadon et al., 2024). Advances in computing technology and the increasing complexity of data have led to the diversification of machine learning algorithms, which are designed to efficiently extract data insights. Typically, these algorithms aim to solve optimization problems using structural risk minimization (SRM), represented as follows.

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Here, *m* denotes the number of training samples. The first term represents the empirical risk, *L*(⋅) is the loss function, is the parameter vector, and the second term, *R*(*f*), is the regularization term accounting for model complexity. The trade-off parameter, ranging from 0 to 1, balances empirical risk and model complexity.

Most research on loss functions focuses on improving or designing them for specific algorithms or problems. A comprehensive understanding of classical loss functions across various fields facilitates the quick selection of suitable loss functions when designing algorithms and inspires improvements to address existing algorithmic challenges. This proposal emphasizes the differences among loss functions in regression, as summarized in Table 1 (Wang et al., 2020).

The square loss function, commonly used in regression problems, measures the squared error between the true value *y* and the predicted value *f*(*x*). Its gradient is variable; it is large for samples with high loss and decreases as the loss approaches zero. However, the square loss is sensitive to outliers because it squares the error *e* =y - *f*(*x*), causing the loss to grow rapidly for large |*e*|. This sensitivity leads the model to focus excessively on outliers, adjusting parameters to minimize their loss. Consequently, the prediction accuracy for normal samples may decrease, reducing overall performance.

The absolute loss function, commonly used in regression, measures the absolute error between the true value *y* and the predicted value *f*(*x*). Unlike the square loss function, it does not increase rapidly with larger errors, making it more resilient to outliers. Therefore, it is preferable in scenarios where outliers significantly impact model learning. However, the absolute loss function is less popular due to its lack of smoothness near zero and its constant gradient regardless of error magnitude, which can reduce optimization efficiency. In this project, the absolute loss function will be linearized to ensure the global optimum. Additionally, minimizing the absolute error in linear regression directly corresponds to using the absolute loss function.

**Table 1.** Loss functions in regression (Wang et al., 2020).

|  |  |
| --- | --- |
| **Square loss** |  |
| Formula |  |
| Algorithm | Linear regression. The objective function is |
| **Absolute loss** |  |
| Formula |  |
| Algorithm | Linear regression. The objective function is |
| **Quantile loss** |  |
| Formula |  |
| Algorithm | Quantile regression. The objective function is |
| **-insensitive loss** |  |
| Formula |  |
| Algorithm | Support vector regression (SVR). The objective function is |

The quantile loss function extends the absolute loss and reduces to absolute loss when the 50th percentile (median) is used. Unlike other loss functions, it allows regression models to provide prediction intervals, offering a range of predicted values rather than a single-point estimate. Quantile loss adjusts the weight of each sample based on the chosen quantile; smaller *τ* penalizes overestimated samples (where the true value is less than the prediction). In contrast, larger *τ* penalizes underestimated samples (where the true value exceeds the prediction). A key application of quantile loss is quantile regression, which is widely used in statistics and econometrics (Koenker & Hallock, 2001). Quantile regression provides interval predictions by varying *τ*, making it suitable for addressing uncertainty and effectively captures data heterogeneity and improves resilience to outliers.

The -insensitive loss function is widely used in support vector regression (SVR) (Drucker et al., 1997). Unlike other loss functions, it does not penalize errors within a given threshold , allowing the model to focus on samples with larger prediction errors. However, because the -insensitive loss includes an absolute term, it is not differentiable but can be linearized. When is zero, it simplifies to the absolute loss function.

Regression is central to statistics. Ordinary least squares regression estimates the conditional mean of the response variable, while least absolute deviation regression estimates the conditional median, offering greater resilience to outliers. Koenker and Bassett (1978) extended the least absolute deviation regression by introducing quantile regression, which estimates the conditional quantile function. Since its inception, quantile regression has become a widely used technique for analyzing the entire conditional distribution of a response variable, providing richer insights beyond the mean or median (Wu & Liu, 2009). Notably, quantile regression includes the least absolute deviation regression as a special case.

Quantile regression, also known for its strong estimation capabilities, has been widely applied in statistical prediction (Yang & Dong, 2019). A quantile represents the value below or above where a certain proportion of data falls. In regression, it estimates the conditional quantile or median of the response variable relative to the predictor variables. For the 50th percentile, quantile loss reduces to the mean absolute error (MAE). Beyond this, it extends MAE by adjusting for different quantiles. Quantile loss makes no assumptions about the parametric distribution of the response variable and provides prediction intervals, making it effective even for residuals with non-constant variance.

Since quantile regression is advantageous for generating interval predictions by given quantile values rather than point estimates, the choice of incorporating quantile determines how positive and negative errors are weighted. Quantile loss can also be applied in neural networks and tree-based models to estimate prediction intervals. We are going to integrate quantile regression into our hybrid method for addressing uncertainty and effectively captures data heterogeneity and improves resilience to outliers.

**1.2 Support vector regression**

SVR, introduced by Vapnik (1995), is a specialized form of regression analysis based on SVM. SVR models the relationship between input features and output values by minimizing prediction errors outside a predefined epsilon margin, while ensuring that observations within this margin are treated as correct, thereby promoting model simplicity and generalization. In contrast, SVM focuses on finding a hyperplane that maximizes the margin between distinct classes. While both involve solving optimization problems, their constraints differ: SVM maximizes the margin between classes, while SVR minimizes the error margin of the function *f*(*x*). Additionally, the loss function in SVR differs from that in SVM (Beniwal et al., 2023). Notably, replacing regression residuals *y-f*(*x*) with 1-*yf*(*x*) can yield a strong loss function for classification tasks.

The introduction of Vapnik's -insensitive loss enabled solving nonlinear regression problems, outperforming conventional techniques (Mukherjee et al., 1997). While multiple linear regression offers simplicity and strong interpretability, it struggles to forecast nonlinear components in load data. Since load series reflect future trends and SVR demonstrates strong empirical performance for nonlinear data (Kim, 2003), integrating SVR can significantly improve forecasting accuracy over linear regression models. SVR has a wide range of applications. For instance, as a non-parametric model, SVR does not require assuming a specific functional form for financial data (Hsu et al., 2009; Patel et al., 2015), thereby reducing errors and effectively filtering noise in financial time series data, which often contain nonlinear patterns and outliers. Unlike conventional models such as autoregressive conditional heteroskedasticity (ARCH), bilinear, and threshold autoregressive model (TAR), which rely on explicit functional forms, SVR captures complex relationships without these assumptions (Hsu et al., 2009; Patel et al., 2015). In electric load forecasting, the use of nonlinear kernel functions in SVR enhances performance by focusing on support vectors, making it less sensitive to outliers and noise (Ceperic et al., 2013; Hahn et al., 2009; Luo et al., 2023).

In addition, Dash et al. (2021) developed an advanced SVR model for time series stock forecasting, demonstrating improved performance and reduced overfitting. Similarly, Luo et al. (2023) proposed a reliable, kernel-free model called weighted quadratic surface support vector regression (WQSSVR) for accurate electric load forecasting, addressing data integrity challenges. Inspired by Luo et al.'s (2023) kernel-free approach, this study seeks to develop a hybrid method, piecewise support vector quantile regression (PSVQR), which simultaneously detects change-points and enhances the explainability of the identified change-points.

**1.3 The framework of the project**

Parametric statistical methods have limitations, including reliance on strict distributional assumptions, which can lead to inaccuracies when these assumptions are unmet. They are often unsuitable for real-time use and struggle to detect complex shifts, such as changes in frequency or auto-correlation. Identifying multiple change-points in multimodal data also remains a challenge. However, non-parametric statistical methods also have several limitations, such as requiring more data to achieve the same precision as parametric methods, being computationally expensive with large datasets, and being prone to overfitting noise due to their high adaptability. Additionally, quantifying uncertainty, such as confidence intervals, can be more challenging in non-parametric settings. Even in hybrid methods, each change-point needs to be examined iteratively (Gupta et al., 2024). Thus, it inspires us to develop a hybrid method to provide prediction interval and detect the change-points simultaneously rather than iteratively.

This project offers four advantages. First, to address uncertainty and provide multiple change-point detection without strict distributional assumptions, the quantile loss function will be used to construct prediction intervals through quantile regression. By leveraging the quantile loss, the conditional quantiles of the response variable can be captured, providing a reliable method for generating prediction intervals that account for non-constant variance and data heterogeneity. Second, it proposes a piecewise model based on a multiple objective programming framework to address structural changes in the data. This approach allows for segmenting the data into distinct regions, where SVR can be applied locally to better model nonlinear patterns and structural changes. The piecewise SVR will be designed to adapt dynamically to variations in the data, ensuring improved resilience against outliers and noise while enhancing prediction accuracy. Third, the number of change points is not limited to only one. Multiple change-points will be detected automatically and simultaneously since the number of change points is formulated as one of the objectives to be minimized. Fourth, the feature selection of input variables and change-point detection will be executed to ensure a parsimonious predicted model. Figure 1 illustrates the characteristics of the proposed method. Overall, this integrated framework—combining quantile regression, multiple objective programming, and piecewise SVR—will provide a more flexible and reliable predictive modeling strategy, particularly for complex datasets with structural changes.

一張含有 文字, 圖表, 字型, 行 的圖片

AI 產生的內容可能不正確。**Figure 1.** The framework of the hybrid model for multiple objective PCD.

**References**

Aminikhanghahi, S., & Cook, D. J. (2017). A survey of methods for time series change point detection. *Knowledge and Information Systems*, *51*(2), 339–367.

Banesh, D., Petersen, M., Wendelberger, J., Ahrens, J., & Hamann, B. (2019). Comparison of piecewise linear change point detection with traditional analytical methods for ocean and climate data. *Environmental Earth Sciences*, *78*, 1–16.

Beniwal, M., Singh, A., & Kumar, N. (2023). Forecasting long-term stock prices of global indices: A forward-validating genetic algorithm optimization approach for support vector regression. *Applied Soft Computing*, *145*, 110566.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Ceperic, E., Ceperic, V., & Baric, A. (2013). A strategy for short-term load forecasting by support vector regression machines. *IEEE Transactions on Power Systems*, 28, 4356–4364.

Chang, W. C., Li, C. L., Yang, Y., & Póczos, B. (2019). Kernel change-point detection with auxiliary deep generative models. arXiv preprint arXiv:1901.06077.

Chatterjee, S., & Qiu, P. (2009). Distribution-free cumulative sum control charts using bootstrap-based control limits, 349-369.

Coello, C. A. C. (2007). *Evolutionary algorithms for solving multi-objective problems*. springer. com.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297.

Csörgö, M., & Horváth, L. (1997). Limit theorems in change-point analysis.

Dash, R. K., Nguyen, T. N., Cengiz, K., & Sharma, A. (2021). Fine-tuned support vector regression model for stock predictions. *Neural Computing and Applications*, *35*, 23295–23309.

Ditthakit, P., Pinthong, S., Salaeh, N., Weekaew, J., Tran, T. T., & Pham, Q. B. (2023). Comparative study of machine learning methods and GR2M model for monthly runoff prediction. *Ain Shams Engineering Journal*, *14*, 101941.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J., Vapnik, V. (1997) Support vector regression machines. In: *Advances in neural information processing systems*, pp 155–161

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, *96*(456), 1348-1360.

Ferger, D. (2018). On the supremum of a Brownian bridge standardized by its maximizing point with applications to statistics. Statistics & Probability Letters, 134, 63-69.

Guo, S., Guo, W., Abolhassani, A., Kalamdani, R., Puchala, S., Januszczak, A., & Jalluri, C. (2019). Manufacturing process monitoring with nonparametric change-point detection in automotive industry. *Journal of Manufacturing Science and Engineering*, *141*(7), Article 071013.

Gupta, M., Wadhvani, R., & Rasool, A. (2024). Comprehensive analysis of change-point dynamics detection in time series data: A review. *Expert Systems with Applications*, 123342.

Gupta, M., Wadhvani, R., & Rasool, A. (2023). A real-time adaptive model for bearing fault classification and remaining useful life estimation using deep neural network. *Knowledge-Based Systems*, *259*, 110070.

Hahn, H., Meyer-Nieberg, S., & Pickl, S. (2009). Electric load forecasting methods: Tools for decision making. *European Journal of Operational Research*, *199*(3), 902-907.

Haque, A. U., Nehrir, M. H., & Mandal, P. (2014). A hybrid intelligent model for deterministic and quantile regression approach for probabilistic wind power forecasting. *IEEE Transactions on Power Systems*, *29*(4), 1663-1672.

Hsu, S. H., Hsieh, J. P. A., Chih, T. C., & Hsu, K. C. (2009). A two-stage architecture for stock price forecasting by integrating self-organizing map and support vector regression. *Expert Systems with Applications*, *36*, 7947-7951.

Jadon, A., Patil, A., & Jadon, S. (2024, January). A comprehensive survey of regression-based loss functions for time series forecasting. In *International Conference on Data Management, Analytics & Innovation* (pp. 117-147). Singapore: Springer Nature Singapore.

Khusna, H., Mashuri, M., Ahsan, M., Suhartono, S., & Prastyo, D. D. (2020). Bootstrap-based maximum multivariate CUSUM control chart. *Quality Technology & Quantitative Management*, *17*(1), 52–74.

Killick, R., Fearnhead, P., & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, *107*(500), 1590-1598.

Kim, K. J. (2003). Financial time series forecasting using support vector machines. Neurocomputing, 55(1-2), 307-319.

Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, 33-50.

Koenker, R., & Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, *15*(4), 143-156.

Kurt, M. N., Yılmaz, Y., & Wang, X. (2020). Real-time nonparametric anomaly detection in high-dimensional settings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(7), 2463–2479.

Lee, J., & Lee, S. (2015). Parameter change test for nonlinear time series models with GARCH type errors. *Journal of the Korean Mathematical Society*, *52*(3), 503-522.

Lee, S. (2020). Location and scale-based CUSUM test with application to autoregressive models. *Journal of Statistical Computation and Simulation*, *90*(13), 2309-2328.

Lee, S., Lee, S., & Moon, M. (2020). Hybrid change point detection for time series via support vector regression and CUSUM method. *Applied Soft Computing*, *89*, 106101.

Lee, S., Na, O., & Na, S. (2003). On the cusum of squares test for variance change in nonstationary and nonparametric time series models. *Annals of the Institute of Statistical Mathematics*, *55*, 467-485.

Lee, S., Tokutsu, Y., & Maekawa, K. (2004). The cusum test for parameter change in regression models with ARCH errors. *Journal of the Japan Statistical Society*, *34*(2), 173-188.

Li, F., & Yang, H. (2024). A novel bounded loss framework for support vector machines. *Neural Networks*, *178*, 106476.

Luo, J., Fang, S. C., Deng, Z., & Guo, X. (2016). Soft quadratic surface support vector machine for binary classification. *Asia-Pacific Journal of Operational Research*, *33*(06), 1650046.

Luo, Jian, Yan, X., & Tian, Y. (2020). Unsupervised quadratic surface support vector machine with application to credit risk assessment. *European Journal of Operational Research*, *280*(3), 1008–1017.

Luo, J., Hong, T., Gao, Z., & Fang, S. C. (2023). A robust support vector regression model for electric load forecasting. *International Journal of Forecasting*, *39*(2), 1005-1020.

Muggeo, V. M., et al. (2008). Segmented: An R package to fit regression models with broken-line relationships. *R News*, *8*(1), 20–25.

Mukherjee, S., Osuna, E., & Girosi, F. (1997). Nonlinear prediction of chaotic time series using support vector machines. *Neural Networks for Signal Processing VII*, 511-520.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, *10*(3), 282-290.

Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, *42*(3/4), 523-527.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, *42*, 259–268.

Pourhabibi, T., Ong, K.-L., Kam, B. H., & Boo, Y. L. (2020). Fraud detection: A systematic literature review of graph-based anomaly detection approaches. *Decision Support Systems*, *133*, Article 113303.

Pushkar, A., Gupta, M., Wadhvani, R., & Gyanchandani, M. (2022, June). A comparative study on change-point detection methods in time series data. In *2022 2nd International Conference on Intelligent Technologies (CONIT)* (pp. 1-7). IEEE.

Sgueglia, A., Di Sorbo, A., Visaggio, C. A., & Canfora, G. (2022). A systematic literature review of IoT time series anomaly detection solutions. *Future Generation Computer Systems*, *134*, 170-186.

Svoboda, R., Kotik, V., & Platos, J. (2024). Data-driven multi-step energy consumption forecasting with complex seasonality patterns and exogenous variables: Model accuracy assessment in change point neighborhoods. *Applied Soft Computing*, *150*, 111099.

Tian, Y., Sun, M., Deng, Z., Luo, J., & Li, Y. (2017). A new fuzzy set and nonkernel SVM approach for mislabeled binary classification with applications. *IEEE Transactions on Fuzzy Systems*, *25*(6), 1536-1545.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267-288.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer, New York.

Wan, C., Wang, J., Lin, J., Song, Y., & Dong, Z. Y. (2017). Nonparametric prediction intervals of wind power via linear programming. *IEEE Transactions on Power Systems*, *33*(1), 1074-1076.

Wan, C., Xu, Z., Pinson, P., Dong, Z. Y., & Wong, K. P. (2013). Probabilistic forecasting of wind power generation using extreme learning machine. *IEEE Transactions on Power Systems*, *29*(3), 1033-1044.

Wang, H., Guo, M., Lee, S., & Chua, C. H. (2022). Forecasting and change point test for nonlinear heteroscedastic time series based on support vector regression. *Plos one*, *17*(12), e0278816.

Wang, Q., Ma, Y., Zhao, K., & Tian, Y. (2020). A comprehensive survey of loss functions in machine learning. *Annals of Data Science*, 1-26.

Wang, X., Borsoi, R. A., Richard, C., & Chen, J. (2023). Change point detection with neural online density-ratio estimator. In ICASSP 2023-2023 IEEE international conference on acoustics, speech and signal processing (pp. 1–5). IEEE.

Warwicker, J. A., & Rebennack, S. (2023). Generating optimal robust continuous piece wise linear regression with outliers through combinatorial benders decomposition. IISE Transactions, 55(8), 755–767.

Wu, Y., & Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica*, 801-817.

Yamada, M., Kimura, A., Naya, F., & Sawada, H. (2013, June). Change-point detection with feature selection in high-dimensional time-series data. In *Twenty-Third International Joint Conference on Artificial Intelligence*.

Yang, L., & Dong, H. (2019). Robust support vector machine with generalized quantile loss for classification and regression. *Applied Soft Computing*, *81*, 105483.

Yu, P. S., Chen, S. T., & Chang, I. F. (2006). Support vector regression for real-time flood stage forecasting. [*Journal of Hydrology*](https://www.researchgate.net/journal/Journal-of-Hydrology-0022-1694?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19), *328*, 704-716.

Yu, J. R., Tzeng, G. H., & Li, H. L. (2001). General fuzzy piecewise regression analysis with automatic change-point detection. *Fuzzy Sets and System*s, *119*, 247-257.

(二) 研究方法、進行步驟及執行進度。請分年列述：1. 本計畫採用之研究方法

與原因及其創新性。2. 預計可能遭遇之困難及解決途徑。3. 重要儀器之配合使

用情形。4. 如為須赴國外或大陸地區研究，請詳述其必要性以及預期效益等。

## Pattern change detection model review

The major PCD models presented by Gupta et al. (2024) are first discussed. Let represent a time series sequence *i*. PCD can be formulated as a hypothesis testing problem: the null hypothesis asserts "No pattern change occurs in the sequence," while the alternative hypothesis 𝐻𝑎 asserts "A pattern change occurs in the sequence.”

Here, *d* represents the threshold value determined by the application's requirements. ​ denotes the starting point, ​ denotes the endpoint, and represents the change-point of the time series. A time series may also contain multiple change-points, represented as , where each element marks a significant change in the sequence where represents the number of pattern change. However, this complexity makes the detecting process more complex while performing multiple change-point detection. Three types of PCD methods are identified: parametric approaches, non-parametric approaches, and hybrid approaches. A brief introduction to these methods is as follows.

**2.1 Statistical methods based on parametric approaches**

**Parametric statistical methods** assume that data follows a predefined distribution (e.g., normal or Poisson) with specific parameters (Aminikhanghahi & Cook, 2017). These methods are effective with well-defined distributions but also have notable limitations, which are discussed in the subsequent sections.

**Models based on piecewise linear regression (PLR)**

Muggeo et al. (2008) and Banesh et al. (2019) proposed a piecewise linear regression method for detecting pattern changes in time series using:

In this method, is the data point at time *i*, with parameters ,, and , representing the change-point . The indicator function equals to 1 when the condition is met and 0 otherwise, while denotes the error term. Piecewise linear regression (PLR) segments a time series into linear relationships, identifying change-points where these relationships shift by minimizing a cost function. Recent advancements, such as those by Warwicker & Rebennack (2023), optimize piecewise linear functions with breakpoints and distance metrics. This approach is effective for anomaly detection, as outliers distort the model, highlighting changes in the data. However, *fn* is generated repeatedly rather than obtained at the same time.

**Models based on cumulative sum (CUSUM)**

Chatterjee and Qiu (2009) proposed a method based on the CUSUM (cumulative sum) approach, where pattern change is calculated using:

In this method, is the cumulative sum at time , ​ is the data point at time, and is the reference value. A change-point is detected when ​ exceeds a threshold , after which the statistic is reset to zero.

Here, *z* controls the false alarm rate, and a change-point (CP) is detected when . The CUSUM test is widely popular for change-point detection in time series due to its simplicity and the extensive literature available on its application. Khusna et al. (2020) improved CUSUM using bootstrapping to estimate the statistic's distribution without strong assumptions. Recent advances by Kurt et al. (2020) focused on detecting rapid deviations in data streams by analyzing summary statistics from high-dimensional data to identify anomalies.

Parametric statistical methods have limitations, including reliance on strict distributional assumptions, which can lead to inaccuracies if unmet. They are often unsuitable for real-time use and struggle with detecting complex shifts, such as changes in frequency or auto-correlation. Identifying multiple change-points in multi-modal data also remains a challenge. The proposed PSVQR can identify multiple change-points at the same time without strict distributional assumptions.

**2.2 Statistical method-based non-parametric approaches**

Non-parametric methods focus on the intrinsic properties of data, using algorithms that operate without specific distributional assumptions. This flexibility allows them to handle diverse data types effectively. These methods excel at detecting structural changes, shifts in median or variance, and non-linear patterns, making them versatile for various applications.

**Models based on unconstrained least squares importance fitting (uLSIF)**

Wang et al. (2023) proposed a uLSIF by optimizing the objectives of the Kullback-Leibler importance estimation procedure (KLIEP). The formula for KLIEP is as follows:

Here, is a *n*-dimensional vector with real-valued components. KLIEP reduces the empirical Kullback-Leibler divergence between true and estimated ratios, while uLSIF minimizes the squared error between them (Yamada et al., 2013). uLSIF optimizes coefficients ​ to reduce the mean squared error between the true ratio and the estimated ratio. Unlike KLIEP, uLSIF is an unconstrained optimization problem, making it more stable and computationally accessible.

**Models based on kernel density estimation**

Chang et al. (2019) proposed a method to detect change-points by calculating kernel density using:

Here, *K* is the kernel function, and *B* is the bandwidth. Kernel change-point detection (KCPD) is a non-parametric method that uses kernel functions to estimate the likelihood density of a dataset. It identifies change-points (CPs) by comparing kernel density estimates before and after potential CPs. A CP is detected when the difference in densities exceeds a specified threshold.

Here, *d* denotes the threshold value determined by the application. KCPD effectively detect shifts in complex, non-parametric distributions and supports diverse kernel functions, such as Gaussian and polynomial, for adaptability.

In PCD, non-parametric statistical methods face several limitations. They often require larger datasets to achieve the same precision as parametric methods and can be computationally intensive with large datasets. PCDs are also susceptible to overfitting due to their high flexibility. Furthermore, quantifying uncertainty, such as constructing confidence intervals, is generally more challenging in non-parametric frameworks (Gupta et al., 2024). To address the issue of uncertainty, we will incorporate quantile regression into our hybrid model to provide interval predictions rather than point estimates.

**2.3 Hybrid methods for detecting pattern changes in time series**

Hybrid methods combine the strengths of various techniques to improve the performance and effectiveness of PCD (Pushkar et al., 2022; Gupta et al., 2024). The main limitation of machine learning algorithms compared to classic statistical methods like ARIMA (Box et al., 2015) is their inability to inherently capture interdependencies among data points, making them less suitable for out-of-the-box forecasting tasks (Svoboda et al., 2024).

**Hybrid approach integrating ARMA and CUSUM**

Lee et al. (2020) proposed a hybrid method, location and scale-based cumulative sum (LSCUSUM) to detect change-points using the ARMA model. The formula for LSCUSUM is as follows.

Here, , are the parameters of the autoregressive and moving average parts, and represents the white noise error terms. SVR, a variant of SVM, is used for regression tasks. SVR aims to identify a function 𝑓() that deviates at most by from the target values , while maintaining smoothness. A significant deviation between actual residuals (*et​*) from the ARMA model and predicted residuals () from the SVR model may indicate a change-point. A change-point is detected if the deviation exceeds a threshold *d*:

where the threshold (*d*) must be given in advance. The choice of *d* depends on the variance of the residuals and the desired sensitivity of the PCD mechanism.

The conventional estimate-based CUSUM test (Lee et al., 2003) struggles with size distortions and low power in complex models. Residual-based CUSUM tests (Lee et al., 2004; Lee & Lee, 2015) address some of these issues but lose power in detecting location parameter changes. Lee (2020) introduced the LSCUSUM test, which leverages both observations and residuals for greater flexibility.

Accurate residual estimation is key to constructing the LSCUSUM test, as residuals reflect prediction errors. While ARMA models are widely used for linear time series, they struggle with nonlinearities. Nonparametric methods, such as recurrent neural networks (RNN) and SVR, address this issue. RNNs effectively handle nonlinear and non-stationary series but face challenges such as overfitting and parameter tuning. In contrast, SVR offers better accuracy, flexibility, and a balance between training and generalization errors, adhering to the structural risk minimization principle. Leveraging these advantages, Lee et al. (2020) used SVR for time series prediction and ARMA residuals to construct the LSCUSUM test for change-point detection.

**LSCUSUM test for ARMA models**

Lee (2020) introduced a CUSUM test for location-scale models (LSCUSUM) relying only on observations and residuals.

Consider the stationary ARMA(*p*, *q*) model as follows:

,

Let and be real numbers, and let be iid random variables with mean zero, covariance , and a finite fourth moment. The objective is to test whether the conditional mean of given past information or the variance of the error terms changes over time *t*. A change occurs if the parameter vector shifts. Based on the observations , the null and alternative hypotheses are formulated as follows:

Under the null hypothesis , ​ remains constant for all *t.*

Under the alternative hypothesis , ​ for some , with for and for , where and represents the change-point.

In implementing the SVR, the coefficients should be appropriately specified. For simplicity, .

**Prediction based on SVR-ARMA model**

Suppose that a training sample is given. Here, and are also used as validation samples. It is assumed that the sample is generated from the ARMA model:

where is an unknown function to be estimated, are nonnegative integers that should be properly determined, and are iid random variables with zero mean and a finite variance. If the training sample is known to follow an SVR–ARMA(*p*, *q*) model with specific orders *p* and *q*, the model can be used. However, if this is not the case, the orders *p* and *q* are determined from the training sample through the following procedure:

**Step 1:**  
Fit a long AR(p) model (*p*→∞, e.g., *p* =log (*T*) using least squares to compute residuals {*et*}.

**Step 2:**  
Fit an SVR–ARMA(*p*, *q*) model to {*et*}. Update residuals recursively using SVR until convergence, yielding {}.

**Step 3:**  
For *p*, *q* ≤ *g*, fit SVR–ARMA models to {}. Compute RMSE for each model and select the optimal (*p*∗, *q*∗) that minimizes RMSE.

**Step 4:**  
Train the SVR–ARMA(*p*∗, *q*∗) model on the full dataset. Use it to compute prediction errors for test samples.

Based on cumulative sums of residuals derived from a fitted model, by analyzing deviations from expected values, the test identifies change-points where the data's underlying structure may have shifted, making it particularly useful in quality control and econometric analysis. Lee et al.'s (2020) residual-based LSCUSUM test is simple and flexible but struggles with delayed detection near change-points. Wang et al. (2022) addressed this limitation by incorporating a time-varying upper control limit and Ferger's (2018) test statistic, enhancing detection efficiency. While the LSCUSUM test is retrospective, it is particularly effective for quickly identifying change-points near the current observation, making it suitable for dynamic monitoring and timely model updates, especially when applied with a moving window scheme (Wang et al., 2022). However, despite these improvements, Lee et al.'s (2022) method still requires multiple iterative steps to identify all change-points. To overcome this, a piecewise quantile SVR with automatic change-point detection will be developed to identify change-points in a single step. We will include the CUSUM (Chatterjee and Qiu, 2009) and LSCUSUM test methods (Lee et al., 2020) as benchmark models for detecting pattern changes or change-points.

## 2.4 Support vector regression model

SVR optimizes a line that minimizes errors within the epsilon tube region, as shown in Figure 2. Using the -insensitive loss function, SVR tolerates acceptable absolute errors, with data points inside the epsilon tube incurring zero error, while those outside contribute to the overall error (Ditthakit et al., 2023).

一張含有 鉸鏈 的圖片

AI 產生的內容可能不正確。

**Figure 2.** Linear SVR with -insensitive loss function (Yu et al., 2006).

To account for data errors, slack variables and , along with a penalty term *C*, are introduced. These slack variables allow for errors outside the margin, , referred to as the soft margin (Cortes & Vapnik, 1995). The regression function is derived by solving the following optimization problem:

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |

The parameter *C* balances the trade-off between the complexity of *f*(*x*) and prediction accuracy by adjusting the emphasis on deviations within a given tolerance. SVR typically employs the -insensitive loss function, defined as:

The -insensitive loss function, defined for a user-specified, , is resilient to outliers and less sensitive to noise. The penalty constant *C* and can be chosen based on experience or determined through cross-validation. The resulting support vector function is *f*(*x*)=. SVR constructs a tube around the regression line with a radius of , where predictions within the tube do not incur a penalty in the loss function. The SVR approach can be enhanced with kernel techniques like the radial basis function (RBF) kernel and polynomial (Poly) kernel. These kernel methods enable modeling nonlinear relationships in data by transforming them into a linear form in the corresponding feature space.

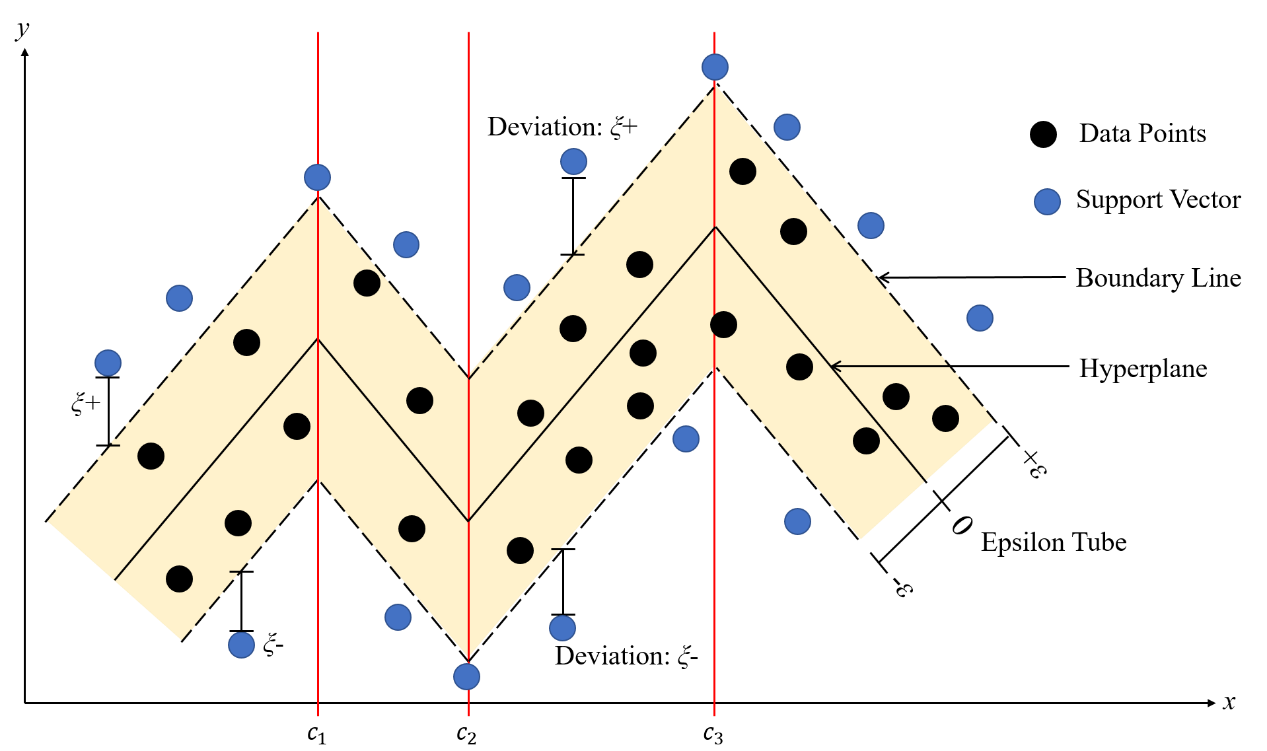
**2.5 The proposed model**

Quantile regression (Koenker and Bassett, 1978) has become an essential vital tool for analyzing the entire conditional distribution of a response variable, and it has been widely used in applied statistics over the past three decades. Wu and Lin (2009) highlighted the importance of variable selection in penalized quantile regression by introducing the adaptive LASSO method. Conventionally, prediction intervals (PIs) are constructed using parametric assumptions (Wan et al., 2013) or quantile error analysis (Haque et al., 2014). Wan et al. (2017) proposed a machine learning-based linear programming (MLLP) model that combines extreme learning machines and quantile regression to generate high-confidence nonparametric PIs for wind power. They also introduced a sensitivity analysis framework for probability mass bias to adaptively optimize quantile proportions. This MLLP-based PI construction is one of key components of the proposed model to provide the interval prediction under uncertainty.

On the other side, straight lines often fail to effectively separate samples in practical applications due to their nonlinear nature. As a result, piecewise linear segments, curves, or polygonal regions are more suitable, although more complex to implement. Figure 3 illustrates piecewise support vector regression using the -insensitive loss function. Achieving high accuracy with SVR often requires substantial computational effort to select an appropriate kernel function and its parameters. Additionally, a singular kernel matrix, common in large load forecasting datasets, can significantly impact forecast accuracy and computational efficiency. To address these challenges, kernel-free SVM models, as proposed by Luo et al. (2016, 2020, 2023) and Tian et al. (2017), utilize quadratic surfaces for classification without relying on kernels. Instead of relying on kernels to project data into higher-dimensional space, this study aims to develop a kernel-free model for improving explainability especially on the detected change-points, leveraging piecewise regression.

The piecewise SVR with quantile losswill be integrated while simultaneously doing feature selection and change-point detection. This allows for flexibility in adjusting the error tolerance , removing the redundant features, offering prediction intervals with various quantiles and incorporating the advantages of change-point detection as a mathematical programming model.

This project will employ piecewise regression (Yu et al., 2001) rather than linear regression and automatically detect the change-point. This approach enables better fitting while simultaneously capturing structural changes. Therefore, piecewise support vector quantile regression with automatic change-point detection will be proposed (PSVQR) by leveraging the strengths of piecewise regression, quantile regression and the SVR.



**Figure 3**. The concept of piecewise support vector regression.

**Multivariate piecewise linear regression**

Assume there are *n* variables and *m* training data. The output is generally related to the *n* input variables. According to the raw data, the *kj* different values for the *jth* input variable are considered. Therefore, in the beginning, *x*1, *x*2,…, *xn* have *k*1, *k*2,…, *kn* change-points, respectively. Every data point is treated as a change-point, and the predicted function is initially presented as follows:

= *a*+

To provide an interval prediction with simultaneous change-point detection, a piecewise term is introduced as follows:

If *pt* is a change-point, then the operation of the piecewise term is as follows (Yu et al., 2001):

*bjt* (|*xj* -*pt* |+ *xj* -*pt* )/2 =

where *t* = 1, …, *kj* and for *j*=1, …, *n.* To control the number of change-points, a coefficient rule is defined as follows.

**Coefficient Rule:** The point *pjt* is considered a change-point if the coefficient of the piecewise term, |*bjt*|>, and is a small positive value specified by the user. Otherwise, *pjt* is not considered a change point if |*bjt*|.

Note that the coefficients of the piecewise term for the model are *bjt,* and the mechanism for controlling the number of change-points by considering multiple objectives will be developed. To handle overfitting and uncertainty, a multiple objective approach will be used for prediction interval estimation based on quantile loss and SVR. In multiple objective optimization (MOP), optimal solutions can detect change-points based on the priority of objectives. Unlike global optimization, which seeks a single solution, MOP focuses on finding trade-offs among objectives (Coello et al., 2007). The concept of Pareto dominance is widely used in MOP to compare candidate solutions.

Since the coefficients of *wj* and *bjt* are unrestricted in sign, they can be reformulated and linearized as:

*wj* = and *bjt* =respectively.

=

The above formulation represents a multivariate piecewise regression model. Variable selection is crucial in model building, especially when numerous candidate predictors are initially included to minimize potential modeling bias (Fan & Li, 2001). However, retaining irrelevant predictors complicates model interpretation and may reduce predictive accuracy. In the regularization framework, penalties like the ​ penalty in LASSO (Tibshirani, 1996) are often used for variable selection. Moreover, to improve variable selection more effectively, we set the variable selection mechanism in the objective function. Therefore, four objectives are outlined in Eq. (2.5.1). The first objective is to use the *ℓ*1-norm to minimize; the second objective is to minimize the quantile loss function, , given the quantile ; the third objective is to reduce the proportion of the number of input variables, where *g* is the number of selected input variables; and the fourth objective is to minimize the number of change-points for a parsimonious fitting function, where and represent the initial number of *xj* distinct value and the generated number of change-points for *jth* input variable for *j* =1, …, *n.*. In addition to adopting the regularization of model coefficients and the quantile loss function, this model aims to minimize the use of fewer input variables, remove the redundant variables, and reduce the number of change-points. Thus, this study proposes the PQSVR to detect change-points and achieve better fitting. The goal is to obtain a parsimonious model with fewer attributes. The fewer the feature selection and change-points, the better. These four objectives are integrated into a single objective using a simple weighting method, as shown in Eq. (2.5.1). As a result, the predicted interval under different quantile s and detected change-points will be generated via multiple objective programming. Furthermore, the time series model can be represented as a regression (Jadon et al., 2024), allowing the proposed method to be adapted to solve the pattern change detection problem in time series.

**PQSVR**

(2.5.1)

s.t.

for *i* = 1, …, *m*, (2.5.2)

for *i* = 1, …, *m*, (2.5.3)

, for *j* = 1, …, *n*, (2.5.4)

, for *j* = 1, …, *n*, (2.5.5)

, for *i* = 1, …, *m*, (2.5.6)

, for *j* = 1, …, *n*, *t* = 1, …, *kj*, (2.5.7)

, for *j* = 1, …, *n*, *t* = 1, …, *kj*, (2.5.8)

(2.5.9)

(2.5.10)

for *j* = 1, …, *n, t* = 1, …, *kj*,(2.5.11)

Testing PSVQR 20250909 effeciently control number of the change point

**PQSVR**

(2.5.1)

s.t.

for *i* = 1, …, *m*, (2.5.2)

for *i* = 1, …, *m*, (2.5.3)

, for *j* = 1, …, *n*, *t* = 1, …, *kj*, (2.5.7)

, for *j* = 1, …, *n*, *t* = 1, …, *kj*, (2.5.8)

(2.5.10)

for *j* = 1, …, *n, t* = 1, …, *kj*,(2.5.11)

17737.85554 + -0.0124\*yt-1 + (1)\*(|yt-1 – 17090.4| + yt-1 – 17090.4)/2

Testing PSVQR 20251009 model now

**PQSVR**

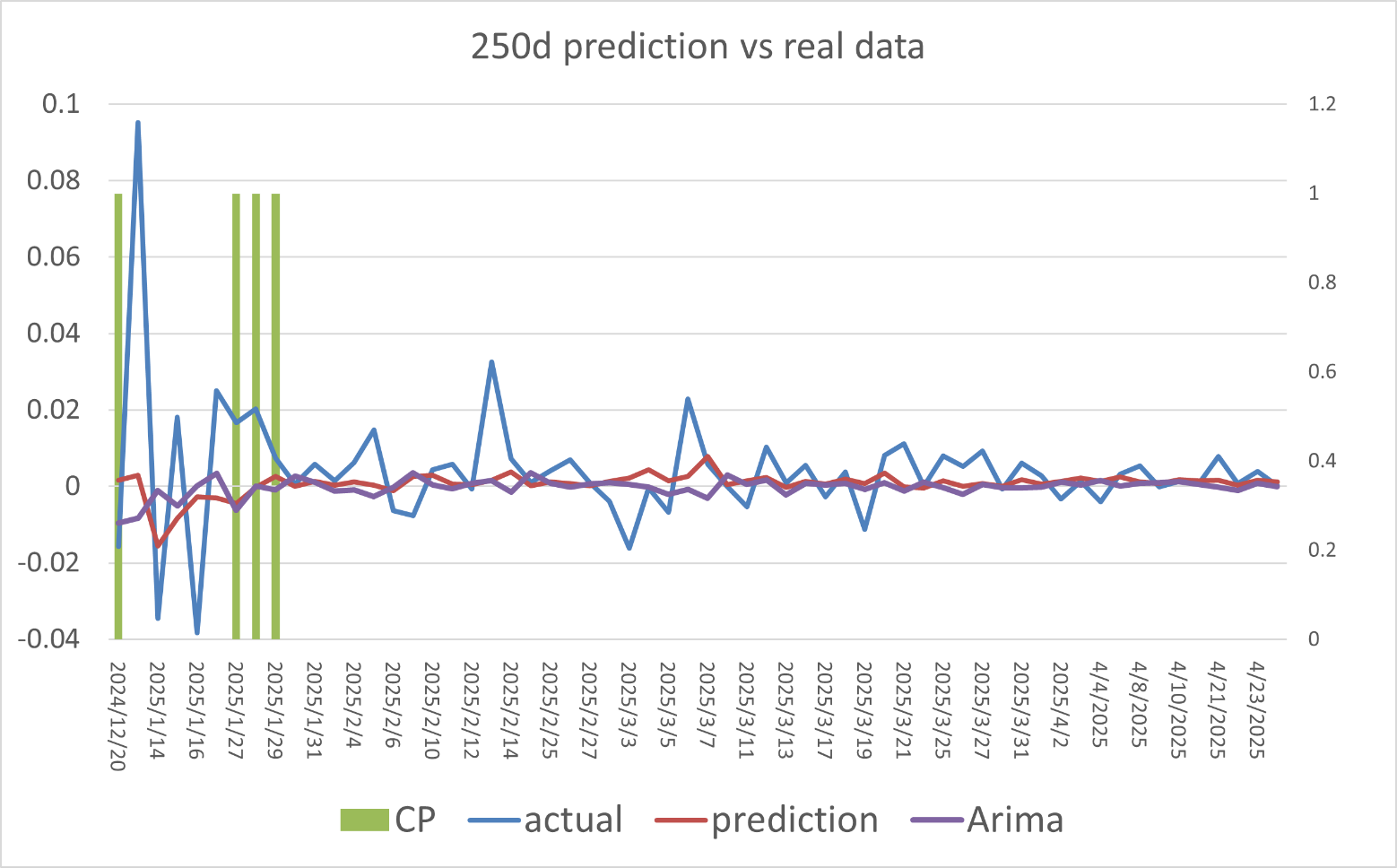
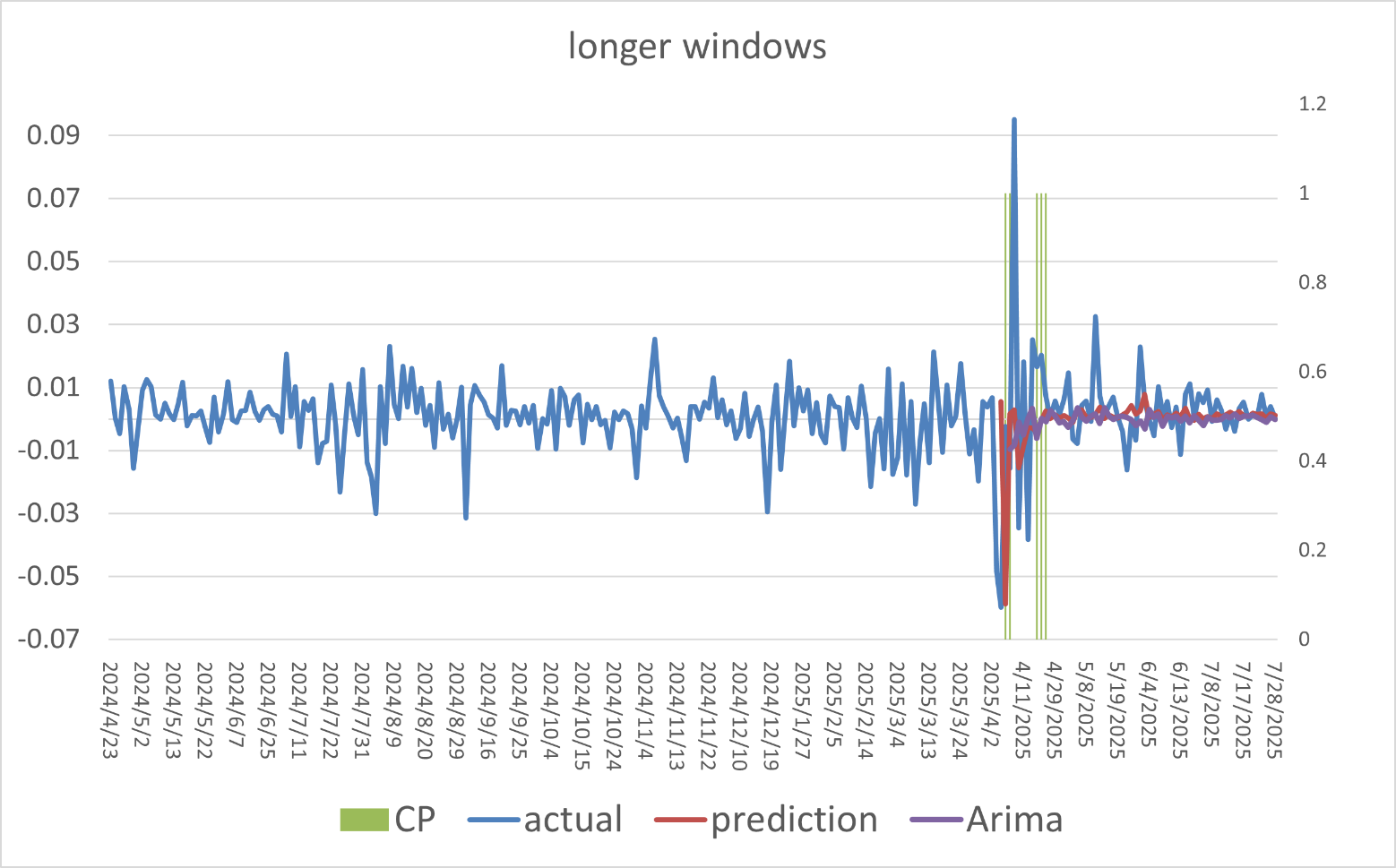
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | |  | **(1)** | |
| *s.t.* | |  |  | |
| 因為當時多 feature 的需要跑很久，所以改成自迴歸的模式，去掉 feature selection 等 | |  | **(2)** | |
|  | |  | **(3)** | |
| 因為 lingo set 的大小無法動態更動，所以加了這個 | , for *j* = 1, …, *n*, *t* = 1, …, *kj*, | | | **(4)** | |
| , | , for *j* = 1, …, *n*, *t* = 1, …, *kj*, | | | **(5)** | |
| , | , for *j* = 1, …, *n*, *t* = 1, …, *kj*, | | | **(6)** | |
|  | , for *j* = 1, …, *n*, *t* = 1, …, *kj*, | | | **(7)** | |
| Cp 至少要 >= 5 個 |  | | | **(8)** | |
|  | , for *j* = 1, …, *n*, *t* = 1, …, *kj*, | | | **(9)** | |
|  | , for *i* = 1, …, m, *j* = 1, …, *n*, *t* = 1, …, *kj*, | | | **(10)** | |

(不用表格會變這樣)

Future work: 導入時間序列，前面影響後面，後面無法影響前面

**測試：**CP 至少5個

設 以完成實驗

****

**一張含有 文字, 行, 繪圖, 螢幕擷取畫面 的圖片

AI 產生的內容可能不正確。**

Math function in the next page

**第一個資訊 (經過測試，如果調高 eta 可以減少 spiky)**

2025/1/24 spiky

為 1/23 所預測 (共 70 CP)，以下列出最後五項

*yi* = 0.9115 *yi*-1+457.1132

+29.2372(|*yi*-1-5949.91 |+ *yi*-1-5949.91)/2

+(-7.6427)(|*yi*-1-5937.34 |+ *yi*-1-5937.34)/2

+1.3175(|*yi*-1-5996.66 |+ *yi*-1-5996.66)/2

+(-147.1)(|*yi*-1-6049.24 |+ *yi*-1-6049.24)/2

+(-38.6564)(|*yi*-1-6086.37 |+ *yi*-1-6086.37)/2

2025/6/30 spiky

為 6/27 所預測 (共 141 CP) )，以下列出最後五項

*yi* = 0 *yi*-1+5446.9

+(-49.368)(|*yi*-1- 6373.45|+ *yi*-1-6373.45)/2 + (-15.3315)(|*yi*-1- 6445.75|+ *yi*-1-6445.75)/2

+1432.769(|*yi*-1- 6466.58|+ *yi*-1-6466.58)/2 + (-1449.68) (|*yi*-1- 6449.79|+ *yi*-1-6449.79)/2

+(-22.0244)(|*yi*-1-6449.14|+ *yi*-1-6449.14)/2

**第二個資訊，用 256 筆資料，one step ahead 做比較 vs arima**

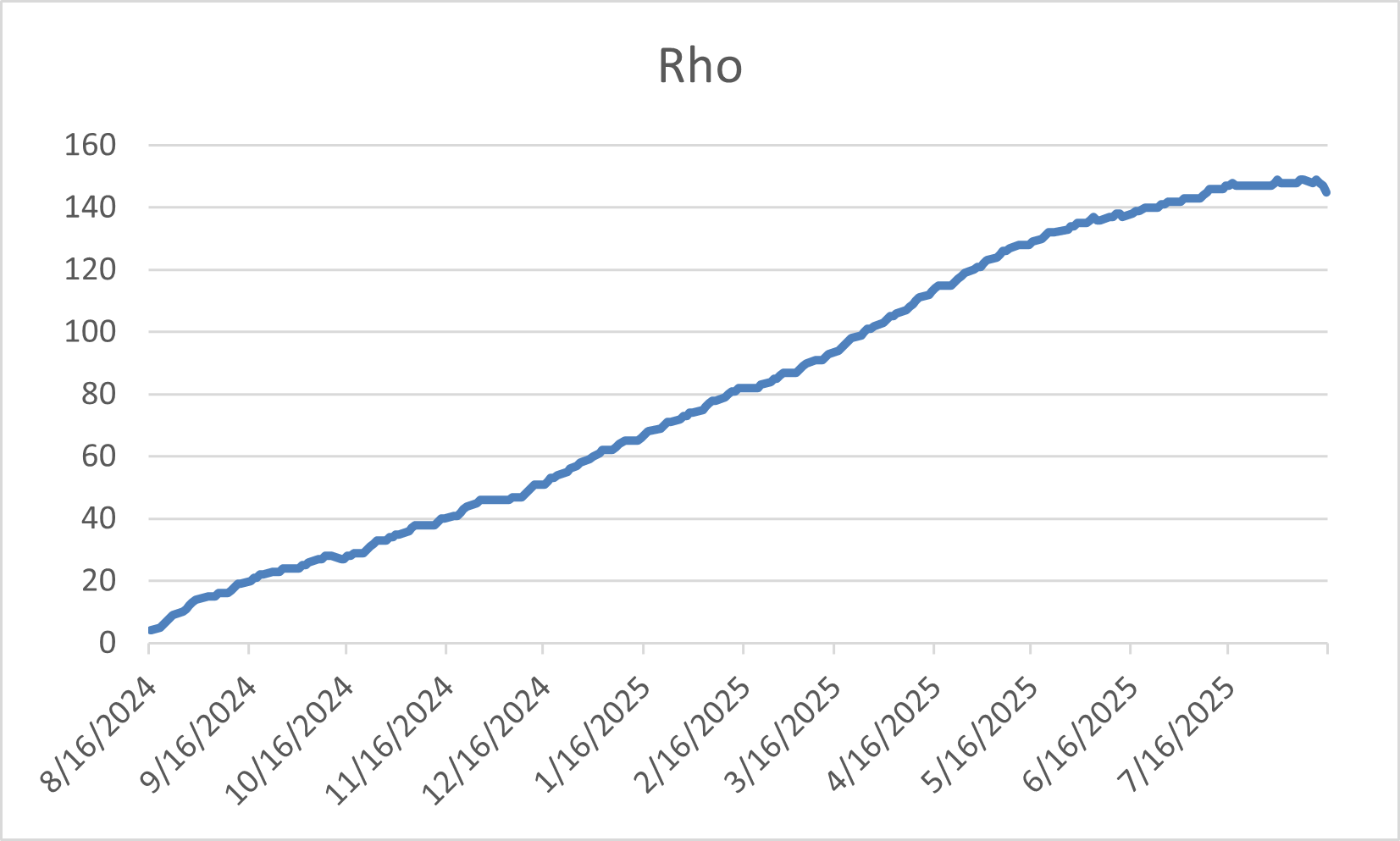
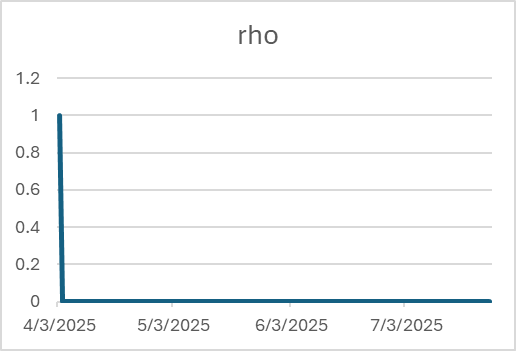
MAE

PSVQR: 0.5374

ARIMA: 0.5819

**第三個資訊，CP數量折線圖 (使用 return 會幾乎沒有 CP，price 則會很多，CP不夠穩定)**

y 為 return



y 為price

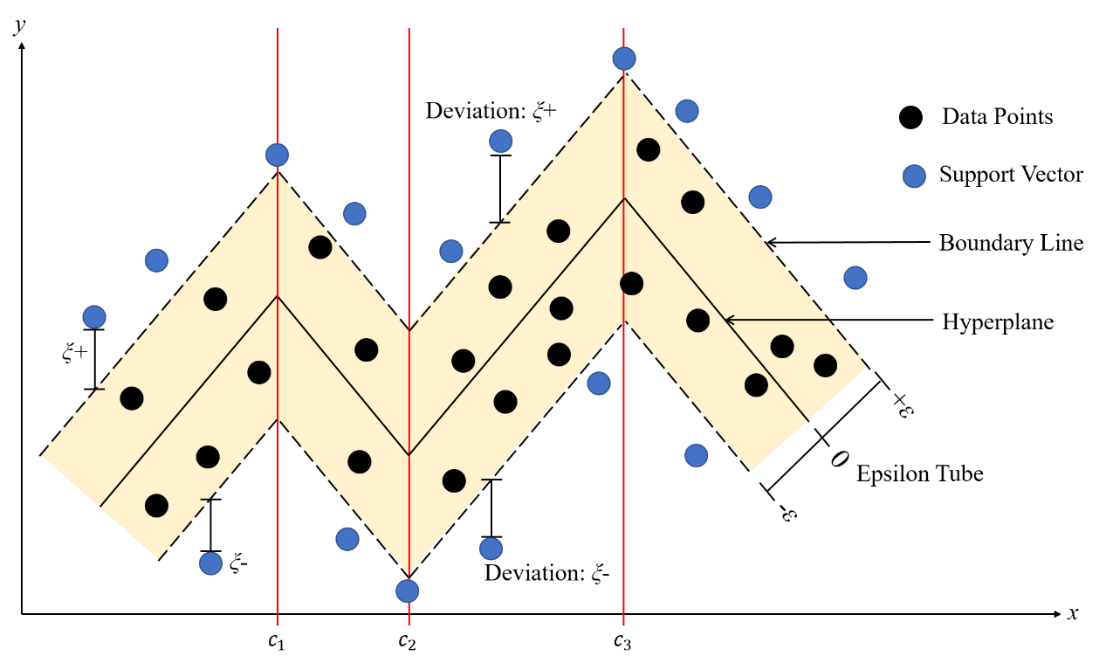
Model 想法：

原本的 svr 在 *min w*，但w實際上很像另一種*b*，所以如果只有*min w* 的話，只會 對第一個 section 做 *minimun* (如下圖 c0 - c1的斜率為 *w*)，但後面的實際斜率則不是 *w*，我猜測有可能是預測不穩定原因。  
而 *b*不是實際上section的斜率，而是當前section斜率減去前一個section的斜率，*bt = St - St-1。*

例如 c1 – c3 的 *b* 會等於c1 – c2 的斜率減去 c0 - c1

所以我覺得要讓預測穩定的話，可以針對每一個 section的slope 做 minimun，general model 在圖下方

*c*0



|  |  |  |  |
| --- | --- | --- | --- |
|  | |  | **(1)** |
| *s.t.* | |  |  |
|  | |  | **(2)** |
|  | |  | **(3)** |
|  | , *t* = 1, …, *n*, | | **(4)** |
|  |  | | **(6)** |
|  | , for *j* = 1, …, *n*, *t* = 1, …, *kj*, | | **(5)** |
| , | , for *j* = 1, …, *n*, *t* = 1, …, *kj*, | | **(6)** |
| , | , for *j* = 1, …, *n*, *t* = 1, …, *kj*, | | **(7)** |
|  | , for *j* = 1, …, *n*, *t* = 1, …, *kj*, | | **(8)** |
|  | , for *j* = 1, …, *n*, *t* = 1, …, *kj*, | | **(9)** |
|  | , for *i* = 1, …, m, *j* = 1, …, *n*, *t* = 1, …, *kj*, | | **(10)** |

where *M* is a large number, is a very small number and is defined by the user. Eqs. (2.5.2) to (2.5.3) show the penalities are taken where are outside. Eqs. (2.5.4) - (2.5.6) and (2.5.11) control the number of input feature. Eqs. (2.5.7) to (2.5.11) remove the redundant suspected change points. The constraint, ensures that a change-point related to the *jth* variable can occur only when the *jth* variable is selected in Eq. (2.5.9). Sensitivity analysis will be performed by varying parameter. For example, the generated model with effective features and change-points will be obtained under various quantile () simultaneously.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Datetime** | **y** | **pre\_1** | **vix** | **future** | | 2025-04-07 16:00:00 | 17418.23 | 17300.34 | 48 | 17388.25 | | 2025-04-08 09:30:00 | 18035.04 | 17418.23 | 38.48 | 17765.25 | | 2025-04-08 10:00:00 | 18084.8 | 18035.04 | 37.09 | 17725 | | 2025-04-08 10:30:00 | 18149.73 | 18084.8 | 37.59 | 17744 | | 2025-04-08 11:00:00 | 17979.43 | 18149.73 | 39.76 | 17785 | | 2025-04-08 11:30:00 | 17843.15 | 17979.43 | 41.36 | 17823 | | 2025-04-08 12:00:00 | 17894.96 | 17843.15 | 40.34 | 18015 | | 2025-04-08 12:30:00 | 17786.14 | 17894.96 | 41.96 | 18005 | | 2025-04-08 13:00:00 | 17525.11 | 17786.14 | 46.11 | 17982 | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

for *i* = 1, …, *m*, (2.5.2)

yi-(

a+pre\_1i\*(w1p-w1n)+vix i ((w2p-w2n)+future i (w3p-w3n)+

(b11p-b11n)( |pre\_1i-17300.34|+ pre\_1i-17300.34)/2+

(b12p-b12n)( |pre\_1i-17418.23|+ pre\_1i-17418.23)/2+

(b13p-b13n)( |pre\_1i-18035.04|+ pre\_1i-18035.04)/2+

(b14p-b14n)( |pre\_1i-18084.8|+ pre\_1i-18084.8)/2+

(b15p-b15n)( |pre\_1i-18149.73|+ pre\_1i-18149.73)/2+

.

.

(b21p-b21n)( |Vixi-48|+ Vixi-48)/2+

(b22p-b22n)( |Vixi-38.48|+ Vixi-38.48)/2+

(b23p-b23n)( |Vixi-37.09|+ Vixi-37.09)/2+

(b24p-b24n)( |Vixi-37.59|+ Vixi-37.59)/2+

(b25p-b25n)( |Vixi-39.76|+ Vixi-39.76)/2+

.

.

(b31p-b31n)( |Futurei-17388.25|+ Futurei-17388.25)/2+

(b32p-b32n)( |Futurei-17765.25|+ Futurei-17765.25)/2+

(b33p-b33n)( |Futurei-17725|+ Futurei-17725)/2+

(b34p-b34n)( |Futurei-17744|+ Futurei-17744)/2+

(b35p-b35n)( |Futurei-17785|+ Futurei-17785)/2+

..)>= (2.5.2)

Eqs. (2.5.2) – (2.5.3) present the initial model as a multivariate regression model with piecewise terms. This model can be extended to a univariate regression or autoregressive framework to support time series analysis. Furthermore, similar to the LSCUSUM approach (Lee et al., 2020), the proposed model incorporates residual fitting with change-point detection.

1. **Experimental design**

A key limitation of machine learning algorithms, compared to traditional statistical methods like ARIMA (Box et al., 2015), is their inability to inherently capture interdependencies among data points, making them less suited for direct forecasting tasks (Svoboda et al., 2024).

This project aims to explore key scientific questions related to algorithms for multiple pattern change detection. This project also examined interpretability, scalability in large data contexts, adaptability to evolving data streams, and the ability to provide actionable insights for timely interventions (Gupta et al., 2024). This study will employ 13 UCI datasets (Li & Yang, 2024) to compare the benchmark models, including the CUSUM and LSCUSUM. Furthermore, the proposed model will be adopted to trace the S&P 500 and NASDAQ 100 indexes to observe structure changes, such as the outbreak of the COVID-19 pandemic or the impact of Trump’s victory in the U.S. election. Since the model is a kernel-free model, this study aims to provide better interpretability of the detected change-points and generate predicted intervals under different quantiles while performing sensitivity analysis. We will also explore the underlying narrative behind the detected change-points.

The optimization and statistical software, including Lingo 18, Gurobi 12.0, and R 4.4.2, will be primarily adopted in this study.

1. **Difficulty and Challenge**

**The issue of a large number of 0-1 variables:** The proposed method involves more binary variables and constraints as the number of input features or samples increase. To address this computational challenge, the use of graphics processing units (GPUs) will be explored to accelerate the computing process if necessary, leveraging advancements such as NVIDIA cuOpt for large-scale linear programming problems (Lu & Yang, 2023, <https://developer.nvidia.com/blog/accelerate-large-linear-programming-problems-with-nvidia-cuopt/?felosearch_translate=1>).

They affirmed the utility of GPUs in solving linear programming by introducing the present cuPDLP.jl, a GPU implementation of the restarted primal-dual hybrid gradient (PDHG) for solving linear programming. It achieves comparable performance to Gurobi, a state-of-the-art LP solver, on standard benchmark sets. This highlights the potential of GPUs and first-order optimization methods. The exploration of parallel computing on linear or mixed integer programming based on the further development of **cuPDLP.jl is of particular interest**. This approach is particularly relevant for handling big data for mixed integer programming problems in the future, as discussed at INFORMS 2024.

1. **Performance measurements and model settings**

To quantitatively compare the performance of predictive models, appropriate metrics must be used to measure the model's fit to the data and its generalization ability. Furthermore, we aim to investigate what occurs at the detected change-points. Four performance measures will be chosen and compared to the benchmark models, such as CUSUM and LSCUSUM.

**Root Mean Square Error (RMSE)**

RMSE comes from which quantifies the overall accuracy of a model and indicates the propagation of error. Lower RMSE values indicate better model performance.

**Mean Absolute Percentage Error (MAPE)**

MAPE is an advancement of and calculates the average absolute percentage error between predicted and actual values. This makes it easy for practitioners to interpret, as it expresses the relative percentage error between predicted and actual values. In finance, relative errors are often more relevant than absolute errors.

Lower MAPE values indicate better model performance.

**Nash-Sutcliffe Efficiency (NSE)**

Initially introduced by Nash and Sutcliffe (1970) to assess the power of models in hydrology, this coefficient is a general form of the coefficient of determination index (*R*2) with values ranging from to 1.

When the model fits perfectly, the value of the coefficient is 1 (NSE = 1). However, when its value is close to , this indicates that the model's predictions are no better than the average data predicted by the model. It penalizes larger prediction errors disproportionately, which may be particularly relevant in the financial field, where downside prediction errors can have significant consequences.

whereandare the actual value, the predicted value at time *i* and the mean of the actual values, respectively.

**Change-Point Neighborhood Error (CPNE)**

A common limitation of these metrics is that they evaluate error over the entire forecast period, potentially overlooking significant changes, such as mean shifts, in the time series. To address this, Svoboda et al. (2024) proposed a parametric change point-related metric called change-point neighborhood error (CPNE), which assesses model accuracy around these critical periods and measures the neighborhood before and after detected change-points.

The CPNE metric is defined by three parameters:

1. : Specifies the neighborhood range before and after the change point event.
2. *S*: A selected change-point detection method, such as binary segmentation, or even manually identified change-points based on external criteria.
3. *E*: A chosen error metric, such as RMSE or MAPE.

CPNE is calculated as the error ratio after and before the change point. A value greater than 1 indicates that the model's accuracy declined following the change-point event.

where

The CPNE will be extended to compare the proposed PSVQR with benchmark models such as CUSUM and LSCUSUM in a comprehensive manner, focusing on the detected change-points that occur before and after these events.

(三) 預期完成之工作項目及成果。請分年列述：1.預期完成之工作項目。2.對於參與之工作人員，預期可獲之訓練。3.預期完成之研究成果（如期刊論文、研討會論文、專書、技術報告、專利或技術移轉等質與量之預期成果）。4.學術研究、國家發展及其他應用方面預期之貢獻。

**We have the following action items:**

**Quantile Regression:**

* Use quantile loss to create prediction intervals for uncertainty and heterogeneity.
* Capture conditional quantiles without strict distribution assumptions.

**Piecewise Model:**

* Develop a PSVQR to handle structural changes and model nonlinear patterns.
* Assess the influence of different loss functions on prediction accuracy.
* Ensure resilience to outliers and improved prediction accuracy.

**Change-Point Detection:**

* Automatically and simultaneously detect multiple change-points by multiple objective programming.

**Feature Selection**

* Perform feature selection for a parsimonious predictive model.

**Framework Testing**

* Validate the integrated method with benchmarks (e.g., CUSUM, LSCUSUM) and complex datasets.
* Uncover the underlying story behind the detected change-points.

**Expected Results and Achievements**

1. To propose the PSVQR model to improve the accuracy of PCD problems and provide prediction intervals for different quantiles.
2. To evaluate the impact of various loss functions on predictions.
3. To analyze the datasets used in the literature.
4. To analyze the S&P 500 or Nasdaq 100 indices and make comparisons.
5. To narrate the story of the detected change-points.
6. To transfer the proposed model from Lingo and Gurobi to cuPDLP.jl, if possible, as this presents the greatest challenge.
7. To present the research outcomes and gather feedback at the INFORMS 2025 Annual Meeting.
8. To support students attending INFORMS 2025, encouraging them to venture beyond their comfort zones and broaden their horizons.
9. To submit at least one paper to reputable international academic journals indexed in SCI or SSCI upon project completion.