Analysing time evolution of density distributions in the financial domain: A literature review and recent technical developments

Yusuke Takase

ABSTRACT

This study investigates the analysis scheme of the dynamics of asset price returns using non-parametric density estimation. The estimation can significantly improve the study of the time evolution of the statistical characteristics of asset price returns using summary statistical approaches, such as variance and skewness, which have been commonly used in financial research for a long time. The study surveys the common features of financial data analytical procedures using non-parametric density estimation and highlights their significance and challenges. In addition, the effectiveness of data analysis techniques that could solve the challenges mentioned above is empirically demonstrated. The study improves the research design and fact-finding for financial time-series dynamic analysis using non-parametric density estimation.

1 Introduction

Understanding asset price dynamics is important in financial econometrics[1] (Campbell et al., 1998). In particular, the development of market systems has made the use of high-frequency trading data[2] available, and understanding the dynamics of high-frequency returns[3] on asset prices has become a major research topic (Barndorff-Nielsen et al., 2009; Brownlees & Gallo, 2006; Engle, 2000). Markets can suddenly crash, as in the financial crisis of 2008 triggered by the collapse of Lehman Brothers and the financial crisis of 2020 induced by the COVID-19 pandemic, inflicting heavy losses on the economy and society. Empirically investigating the statistical characteristics of asset price return dynamics can help develop models and theories of future asset price dynamics, contributing to market

Keywords：High-frequency Financial Data; Kernel Density Estimation; Parallel Processing; Future Distribution Forecasting; Asset Price Dynamics ［受理日 2022 年 10 月 24 日］

participants' rapid risk control and overall market stability.

In financial econometrics, it is common to measure the statistical characteristics of asset price returns in terms of moment statistics[4], such as variance, skewness, and kurtosis (Arnerić, 2020). Consistency with existing financial theory and simplicity of calculation in practice are partial reasons for the widespread use of these statistics. This trend is also the case in high-frequency return dynamics research, in which moment statistics have been actively studied as realised volatility, realised skewness, and realised kurtosis, especially in the field of market microstructure (Amaya et al., 2015; Barndorff-Nielsen et al., 2009; Mei et al., 2017).

However, the conventional summary statistics approach can only partially represent the statistical characteristics of return dynamics. For example, variance indicates the degree of spread of the density distribution, while skewness indicates only the asymmetry of the density distribution. In this respect, it has been seen as problematic over the years that summary statistics fail to make use of the potential shape information on density distribution that should be available from time-series data (Arroyo & Maté, 2009; Bessa et al., 2012; Krempl et al., 2019; Tay, 2015). Indeed, it is known that for time-series data of large size and frequency and non-stationary time-series data[5], approaches using summary statistics hinder a proper understanding of the temporal evolution of statistical features of time-series data (Lampe & Hauser, 2011). In other words, the conventional summary statistics approach in financial econometrics may miss important statistical features and their time-evolution patterns that should be found in asset price returns, especially high-frequency returns.

An alternative approach is to analyse the comprehensive shape of the density distribution of returns obtained directly from the sample and the time evolution of the shape, without making arbitrary assumptions and without summarising the information. The effectiveness of directly analysing the density distribution of returns using non-parametric[6] probability density estimation has been reported (Arnerić, 2020; Iwamoto & Takada, 2018; Semeyutin & O'Neill, 2019; Takada, 2009; X. Wang et al., 2018). This approach is a highly useful time-series analysis method that is increasingly being used interdisciplinary in the time-series analysis domain as well as in the financial domain (Goswami et al., 2018; He & Li, 2018; Kraemer et al., 2021; Krempl et al., 2019; Lampe & Hauser, 2011; Lampert, 2015). However, a common analysis scheme for asset price returns using this non-parametric probability density estimation, the utilisation barriers to be considered, and the effective techniques to solve the barriers have not yet been discussed.

To address these research gaps, this study focuses on the kernel method, commonly used for non-parametric probability density estimation, and reviews the previous literature to explore the procedural features and issues of density distribution analysis of asset price returns. This study makes a unique contribution to the literature because it considers applying non-parametric probability density estimation methods to high-frequency financial data considering recent trends in the financial domain. In addition, an interdisciplinary survey of data analysis techniques that could address the current challenges identified in the review mentioned above is conducted, and the effectiveness of these techniques is tested.

The main contributions of this study are as follows. The first contribution is that, through a comprehensive review of previous studies in the financial domain, common aspects of density distribution studies of returns using non-parametric density estimation are organised, such as its significance, procedures, and points to note. To the best of my knowledge, there is no comprehensive review of existing studies in this context, which is a highly novel endeavour. It provides a better understanding of density distribution analysis of returns using non-parametric density estimation and assists in constructing an analytical design that meets the research objectives. The second contribution is the practical introduction of data analysis techniques that could solve the identified problems of density distribution studies of returns in terms of their effectiveness. It helps to remove barriers to the introduction of non-parametric density estimation methods in studies of higher frequency financial data in general, including studies of returns on price assets.

The rest of this paper is structured as follows. Section 2 provides an overview of non-parametric density estimation. The significance of focusing on density distributions and specific analysis methods are explained. Section 3 presents the results of a previous research review of asset price return studies using kernel density estimation. It highlights the commonalities in the analysis and the issues remaining in the current state of the art. Finally, the effectiveness of data analysis methods that address the issues highlighted in the previous section is tested using synthetic data.

2 Non-parametric kernel density estimation

2.1 Non-parametric density estimation

Non-parametric density estimation is a statistical method for estimating, from sample data, the true density function[7] $f(X)$ that a continuous random variable X follows without assuming a model for the density distribution. Non-parametric density estimation has been applied in a variety of applications and now underpins data analysis techniques, non-parametric regression (Čížek & Sadıkoğlu, 2020; Eubank, 1999; H.-G. Müller, 2012), non-parametric independence tests (Bagnato et al., 2014; Rosenblatt, 1975), and non-parametric pattern recognition (Fukunaga, 2013; Jain et al., 2000). Recent developments in the data analysis environment, such as improved computer performance and increased data volumes, have further expanded the use cases for approaches based on non-parametric density estimation.

There are three general validities of non-parametric density estimation (Izenman, 1991).

A) **Exploratory Analysis.** It aims to explore latent features behind the data in an exploratory manner. Non-parametric density estimation allows for flexible numerical descriptions of features of random variables, such as multimodality[8)] and fat tails[9)].

B) **Confirmatory Analysis.** It aims to test models and hypotheses from the sample data. Non-parametric density estimates are the basis for decision-useful statistical analysis methods, such as independence tests.

C) **Presentation (Visualisation).** With non-parametric density estimation, the statistical peculiarities of the data can be easily explained and interpreted through a graph of density function curves.

Various approaches have been proposed for non-parametric density estimation, including the histogram algorithm (Scott, 1979), frequency polygons algorithm (Scott, 1985), kernel algorithm (Silverman, 1986), spline algorithm (Eilers & Marx, 1996), projection pursuit algorithm (Friedman & Tukey, 1974), and orthogonal series algorithm (Efromovich, 2010). Recent advances in machine learning technology have also led to using non-parametric density estimation approaches with neural network algorithms (Q. Liu et al., 2021; Sindagi & Patel, 2018). Among them, the long-established kernel algorithm is still one of the most popular approaches (Izenman, 1991).

2.2 Kernel method

This study focuses on kernel algorithms and describes non-parametric kernel density estimation computation. Given the data $x = \{X_1, \ldots, X_n\}$, the traditional fixed kernel density estimation $\tilde{f}$ is as follows.

$$\tilde{f} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x-X_i}{h}\right).$$

$K(\cdot)$ is a kernel function[10], where h is a fixed bandwidth that controls smoothness.

When multimodal or fat-tailed distributions are assumed in probability density time series, or when the presence of outliers[11] is assumed in the data, adaptive kernel density estimation robust to those cases is preferred (Iwamoto & Takada, 2018; Takada, 2009). The adaptive kernel density estimation method locally varies the bandwidth of the fixed kernel density estimation (Abramson, 1982; Breiman et al., 1977). Its definition f is as follows.

$$f = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h\lambda_i}K\left(\frac{x-X_i}{h\lambda_i}\right).$$

λ_i are weights that vary the bandwidth locally. The weights λ_i are calculated as $\{\tilde{f}(X_i)/g\}^{-1/2}$ using the pilot estimate[12] from fixed kernel density estimation, where $\log g = n^{-1}\sum\log\tilde{f}(X_i)$. Compared to other major non-parametric probability density estimation methods, the adaptive kernel density estimation changes the bandwidth locally depending on the density of the data, thus accurately capturing non-normal probability density distribution[13] shapes, such as a tail with little data and a peak of gathered data (Takada, 2008).

2.3 Kernel selection

As mentioned above, kernel density estimation requires several parameters: the choice of kernel function $K(\cdot)$ and bandwidth h. Depending on the choice of parameters, the estimated probability density can vary significantly (Arnerić, 2020; Takada, 2008). The choice of parameters in kernel density estimation is, therefore, a very important procedure, as it determines the quality of the probability density estimation results.

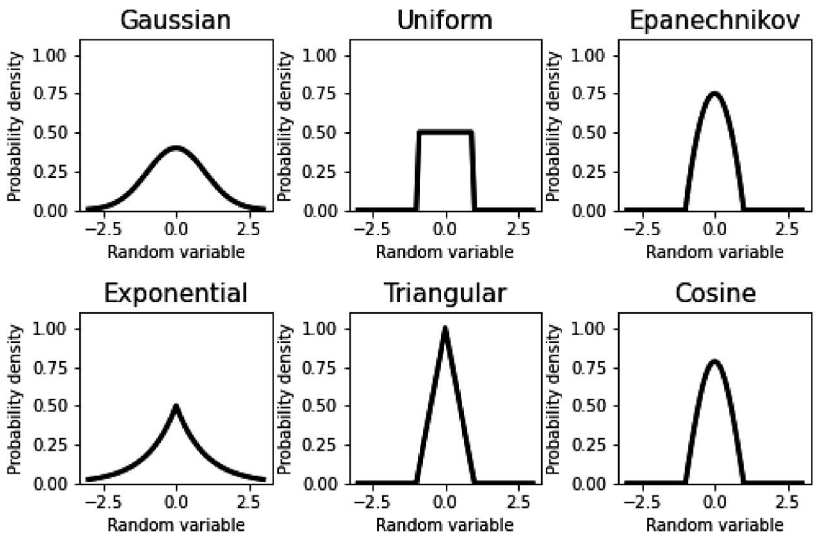
First, the choice of kernel function is explained. Table 1 presents a list of typical kernel functions. Figure 1 shows the kernel functions shown in the same coordinates. The shape of the probability density distribution differs depending on the kernel function. The histogram can be seen as a kernel density estimation method with uniform kernel functions.

The Gaussian kernel is the most common kernel function. It is known that differences in kernel functions have little effect on the results of density estimation (Marron & Nolan, 1988; Scott, 2015; Wasserman, 2006). Therefore, most literature ignores this topic (Y.-C. Chen, 2017). The most natural choice is the Gaussian kernel (Janssen et al., 1995; Silverman, 1986).

Table 1: List of typical kernel functions.

Function name	Function expression
Gaussian	$K(t) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)t^2}$
Uniform	$K(t) = \{ \begin{matrix} 1/2, & \lvert t \rvert \leq 0 \\ 0, & \lvert t \rvert \geq 0 \end{matrix}$
Epanechnikov	$K(t) = \{ \begin{matrix} 3(1-t^2)/4, & \lvert t \rvert \leq 0 \\ 0, & \lvert t \rvert \geq 0 \end{matrix}$
Exponential	$K(t) = \frac{1}{2} e^{-\lvert t \rvert}$
Triangular	$K(t) = \{ \begin{matrix} 1-\lvert t \rvert, & \lvert t \rvert \leq 0 \\ 0, & \lvert t \rvert \geq 0 \end{matrix}$
Cosine	$K(t) = \{ \begin{matrix} \pi \cos(t\pi/2)/4, & \lvert t \rvert \leq 0 \\ 0, & \lvert t \rvert \geq 0 \end{matrix}$

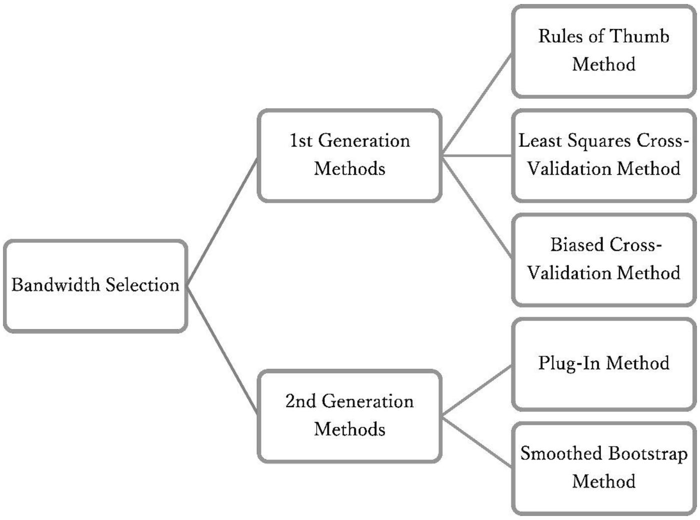
Figure 1: Density distribution shapes of typical kernel functions.



2.4 Bandwidth selection

Next, the bandwidth selection methods are described. As shown in Figure 2, bandwidth selection methods can be broadly classified (Jones et al., 1996) into first- and second-generation types. First-generation bandwidth selection methods include the rule-of-thumb method (Scott, 2015; Silverman, 1986), the least-squares cross-validation method

Figure 2: Representative bandwidth selection methods (Jones et al., 1996)

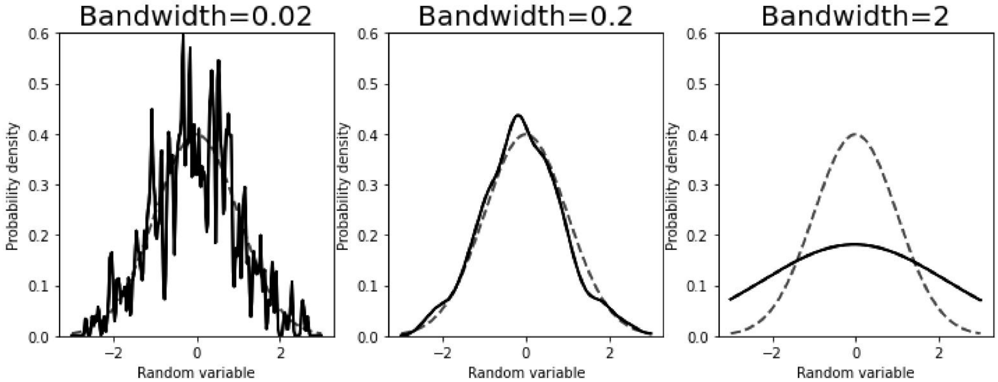


(Bowman, 1984; Rudemo, 1982), and the biased cross-validation method (Scott & Terrell, 1987). Second-generation bandwidth selection methods include the plug-in method (Sheather & Jones, 1991; Woodroofe, 1970) and smoothed bootstrap method (Faraway & Jhun, 1990; Taylor, 1989). The differences between the first- and second-generation methods are classified according to when they were proposed. The details of bandwidth selection methods and performance comparisons are not discussed in this paper, as there are already several useful existing studies (Jones et al., 1996; Scott, 2015; Sheather, 2004; Wasserman, 2006).

Both bias and variance in density estimation results depend on bandwidth (Arnerić, 2020). Figure 3 shows the density estimation results when various bandwidths are applied — the shape of the distribution leads to very different results for different bandwidths. When the bandwidth is too small (left-hand side of Figure 3), local features are overestimated, and the density estimates vary significantly from one random variable to another. If the bandwidth is too large (right-hand diagram in Figure 3), smooth estimation results can be obtained, and the results are largely erroneous. Therefore, using the appropriate bandwidth selection method and estimating the appropriate bandwidth is a very important task in density estimation.

However, all bandwidth selection methods have advantages and disadvantages (Arnerić, 2020). It is essential to understand each bandwidth selection method's advantages and disadvantages and use the appropriate bandwidth selection method according to the

Figure 3: Density estimation results with different bandwidths.



Note: Fixed kernel density estimation was carried out using 300 samples taken from the standard normal random distribution. The grey dashed line shows the theoretical standard normal distribution, and the black solid line shows the estimated probability density distribution. The figure shows that the same density estimation method gives very different results for different bandwidths.

research objectives, data characteristics, and other analysis requirements.

3 Application of kernel density estimation methods to financial time series

3.1 Literature selection criteria

This subsection surveys the literature that applies univariate kernel density estimation, focusing on return studies in the stock market (including stock derivatives[14)]), a particularly representative research subject in the financial time-series research area. This subsection describes the selection criteria.

(a) First, this study investigates papers in the ‘economics, econometrics, and finance’ field of Scopus for which the search terms’ kernel density estimation’, ‘stock’, and ‘return’ are used simultaneously in any search items (e.g. article title or abstract).

(b) Second, to focus only on English-language papers that had already been published, papers that also fulfilled the following criteria were included; literature type ‘Article’, publication stage ‘Final’, publication type ‘Journal’, and text language ‘English’.

(c) Third, the search excluded literature with low citation counts, literature published before 2019 and with zero citations was excluded from the search. By applying conditions (a) to (c) above, the number of eligible references was 83.

(d) Fourth, because conditions (a) to (c) also search for papers that do not apply kernel density estimation to stock market returns (including stock derivatives), the references that apply kernel density estimation to stock market returns were manually

Table 2: List of literature reviewed.

	Configuration of kernel density estimation					Dataset	
	Usage	Algorithm	Kernel	Bandwidth selection	Language (Package)	Object	Scale
Tsay (2016)	Visualisation	Fixed	Gaussian	Rule of thumb method	R (default)	US Index	Daily
Gu et al. (2018)	Visualisation	Fixed	-	plug-in method	-	US index future	Intraday
X. Wang et al. (2018)	Proposal for a new algorithm	adaptive	Gaussian and others	machine learning method	-	US Index	Daily
Cai et al. (2018)	Visualization	Fixed	-	-	MATLAB	US Index	Intraday
Semeyutin and O'Neill (2019)	Proposal for a new algorithm	adaptive	Gaussian and others	least-squares cross-validation method	R	Japan, Europe, US Indices	Daily
Gurrib et al. (2020)	Visualisation	Fixed	Gaussian	Rule of thumb method	Python (Scikit-learn)	US Index	Daily
	Visualisation	Fixed	Gaussian	-	-	US Index	Daily

Note: For each item, the most common choices are underlined. Items not mentioned by the preceding literature are marked with '-'.

extracted from the 83 eligible references.

As a result, the review in this study focuses on a total of seven research papers that satisfy all conditions.

3.2 Literature review

This subsection summarises commonalities in the previous literature that has conducted kernel density estimation analyses on stock market returns (including stock derivatives), particularly concerning their purpose of introducing kernel density estimation and analytical design. Table briefly presents the results of the literature review. For each item, characteristic common terms are underlined.

First, the purpose of introducing kernel density estimation into analysing stock market returns is summarised. The most common purpose is visualising the return distribution (see 'Usage' section in Table). By enjoying the effectiveness of A and C mentioned in Subsection 2.1, it is possible to compare and evaluate the different shapes of the return distribution in a flexible manner. For example, the temporal evolution of return distributions (Cai et al., 2018; Gu et al., 2018; Tsay, 2016) and differences in return distributions by data type and condition (J. Chen et al., 2022; Gurrib et al., 2020; Tsay, 2016) have been identified. Other literature also aims to develop kernel density estimation methods suitable for time-series analysis, such as stock market returns (Semeyutin & O'Neill, 2019; X. Wang et al., 2018).

Second, the analytical design of introducing kernel density estimation into analysing stock market returns is organised. As a result, some commonalities can be identified ('Algorithm' and 'Kernel' entries in Table). The traditional fixed kernel density estimation method is often used as the basic algorithm for kernel density estimation methods. Some literature also combines fixed kernel density estimation with sliding window[15] methods to achieve density estimation that gives importance to the data nearest to the prediction point (Cai et al., 2018; Tsay, 2016). The ease of use by users without information engineering expertise may be one reason for the widespread use of fixed kernel density estimation. Several programming languages already use kernel density estimation by default or have fixed kernel density estimation pre-installed in well-known packages, such as R, Python, and MATLAB[16] (see 'language' entry in Table).

As mentioned above, other literature has shown that exponential weighting adaptively reduces the impact of past data and adaptively emphasises the impact of data closer to the prediction time point by using a new time-adaptive dynamic kernel density estimation (Semeyutin & O'Neill, 2019; X. Wang et al., 2018). However, this approach is still new. There is room for further discussion as to whether it is appropriate for financial time-series analysis with any objectives, such as increasing computational costs or the appropriateness of exponential weighting.

Except for the references that do not mention kernel functions, all references commonly use the Gaussian kernel as the kernel function. As mentioned in Subsection 2.3, it is assumed that the use of the Gaussian kernel is the empirical default in the financial domain as well since the effect of the kernel function on the estimation results is small.

However, there is no commonality regarding the bandwidth selection method (see Table under 'Bandwidth selection'). The rule of thumb method is used most often, but due to the small number of survey samples, it cannot be said that there are clear commonalities. As mentioned in Subsection 2.4, all bandwidth selection methods have advantages and disadvantages (Arnerić, 2020). It is assumed that the results from each study select the appropriate band selection method according to the research objectives, data characteristics, and other analysis requirements.

Third, the datasets covered by the previous literature are summarised. The analysis of stock market returns using kernel density estimation targets US market data with a timescale of days or less (see 'Dataset' in Table). Non-parametric statistical methods, including kernel density estimation, require a certain amount of data to ensure estimation accuracy, making applying them to weekly or monthly data difficult. The availability and importance of intraday[17] data are increasing as market trading, including algorithmic and high-frequency trading[18], is becoming faster. Analysis using intraday data is likely to develop further in the future.

3.3 Future challenges in the financial high-frequency time-series research domain

This section refers to the challenges of applying kernel density estimation methods in the financial domain. The challenge is adapting to the modern financial data analysis environment, especially larger and higher frequency data. With the development of information technology, market trading is becoming faster through algorithmic and high-frequency trading. The analytical significance of investigating the statistical properties of the intraday variability of financial time series is increasing yearly. The analytical effectiveness of high-frequency data in the micro-market structure domain is already known (Falkenberry, 2002; U. Müller, 2001; Verousis & Gwilym, 2010). Kernel density estimation using intraday data, as in the previous literature (Falkenberry, 2002; U. Müller, 2001; Verousis & Gwilym, 2010) presented in the previous section, is likely to become more popular.

There are several barriers to overcome when applying kernel density estimation to high-frequency financial data;

A) the presence of outliers,

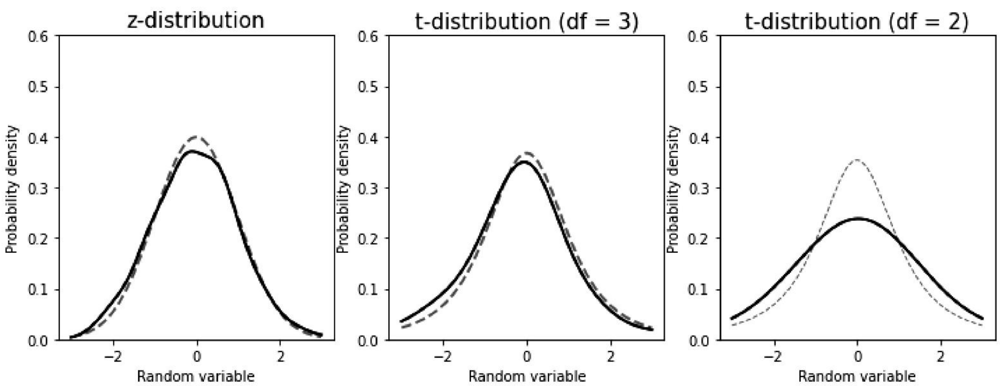
B) high computational costs and

C) difficulties in visualisation.

The behaviour of the tail of the density distribution is of great importance in the intraday analysis of financial time series. The tail of the distribution implies the occurrence of extreme fluctuations (tail events), such as financial crises and crashes, which are dangerous and important signals in the market. It is known empirically that the probability of tail events increases in intraday analysis and takes a fat-tailed density distribution shape (B. Liu et al., 2019). The accuracy of estimating the tail of the density distribution is, therefore, highly significant.

However, outliers exist in high-frequency financial time series (Falkenberry, 2002; U. Müller, 2001; Verousis & Gwilym, 2010). Outliers reduce the reliability of the results of density distribution tail estimation (Barrier A). However, the criteria for outliers are difficult to define, and outlier pre-processing[19] can distort reality if carried out excessively (Falkenberry, 2002; U. Müller, 2001; Verousis & Gwilym, 2010). Therefore, when apply-

Figure 4: Fixed kernel density estimation of fat-tailed distributions.



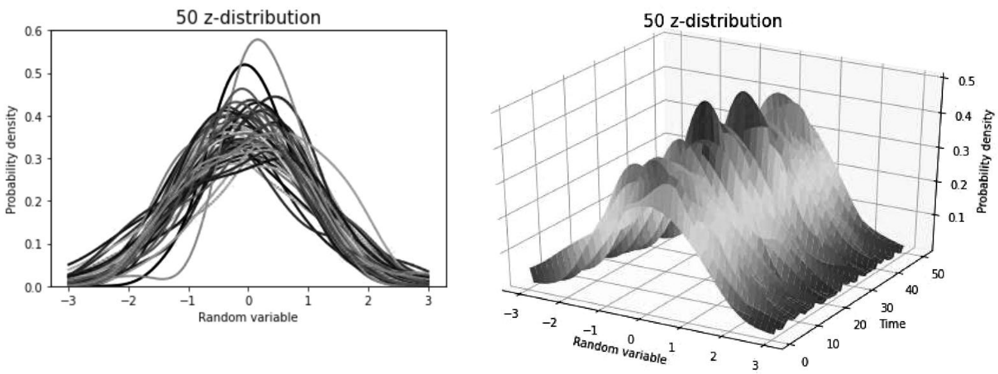
Note: Fixed kernel density estimation was applied to 500 random numbers from the standard normal distribution (left panel), the t-distribution with three degrees of freedom (centre panel) and the t-distribution with two degrees of freedom (right panel). The Gaussian kernel and Sil verman's rule of thumb, which were highly used in the previous literature review, were applied. It is observed that the errors in the estimation results are larger in the more fat-tailed case (right panel).

ing kernel density estimation to high-frequency financial time series, it is necessary to consider using approaches that are robust to outliers and can accurately estimate the tail of the distribution.

Fixed kernel density estimation, the most used in the previous literature (previous section), is vulnerable to noise and, as discussed above in Subsection 2.2, is generally known to have low estimation accuracy in the tail. Figure 4 shows the fixed kernel density estimation results for the fat-tailed distribution. It is observed that the errors in the estimation results are larger in the more fat-tailed case (right panel of Figure 4). In addition, comparative studies have reported that adaptive kernel density estimation has better accuracy for multimodal and fat-tailed distributions than fixed kernel density estimation (Takada, 2008). However, it is assumed that the computational cost of the high accuracy density estimation for the tail of distribution increases and the practicality decreases (barrier B). The impact is even greater when the data is large scale. Kernel density estimation methods suitable for intraday data need to be designed.

In addition, as financial time series become more frequent, more detailed visualisation of the time evolution of density distributions will be required, making it more difficult

Figure 5: Visualisation of multiple density distributions.



Note: Fixed kernel density estimation was applied 50 times to 50 random numbers based on the standard normal distribution and visualised simultaneously. The Gaussian kernel and Silverman's rule of thumb, which were highly used in the previous literature review, were applied. The left-hand figure shows the results visualised in two dimensions, while the right-hand figure shows the results visualised in three dimensions. In the left-hand figure, the differences in the characteristics of each distribution can be seen in detail. However, when there are many density distributions to be visualised simultaneously, the density distribution close to the back cannot be correctly evaluated due to excessive overlap of the figures. In the right-hand diagram, all density distributions can be checked at once. However, if there are many density distributions to be visualised at the same time, the detailed differences in each density distribution cannot be correctly evaluated.

to express using existing visualisation methods (Barrier C). Figure 5 shows the results of visualising many density distributions simultaneously in a single graph. The left hand of Figure 5 shows the results in two dimensions, while the right hand of Figure 5 shows the results in three dimensions. In the left-hand Figure 5, the differences in the characteristics of each distribution can be seen in detail. However, when many density distributions are visualised simultaneously, the density distribution close to the back cannot be correctly evaluated due to excessive figures overlap. In the right-hand Figure 5, all density distributions can be checked at once. However, if many density distributions are visualised simultaneously, each minor difference in the distributions cannot be correctly evaluated. New visualisation expressions suitable for describing the time evolution of high-frequency time-series data will be required in the financial domain.

4 Recent progress in density estimation using kernel methods

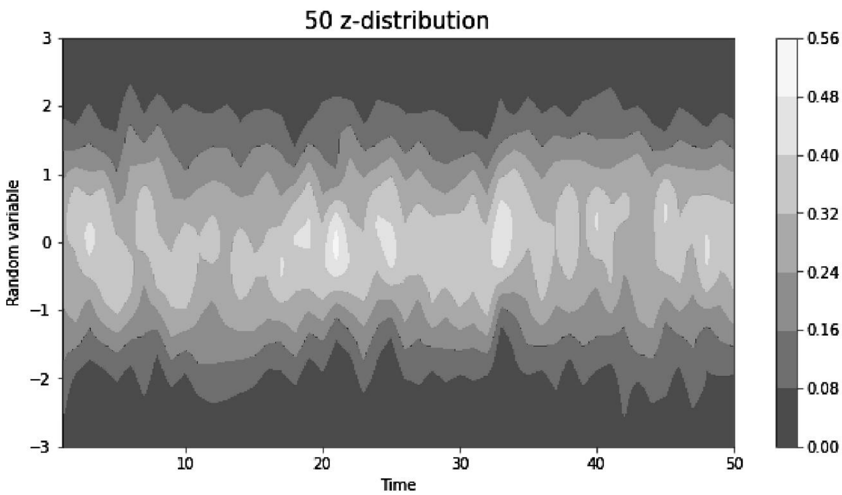
This section examines the techniques that effectively solve future problems identified in the previous section in financial time-series research using kernel density estimation. Kernel density estimation is being used in several time-series analysis domains other than the financial domain; energetics (Bessa et al., 2012; He & Li, 2018), meteorology (Lampe & Hauser, 2011), land and water sciences (Kraemer et al., 2021; R. Wang et al., 2012), oceanography (Goswami et al., 2018), acoustics (Lampe & Hauser, 2011), mechanical engineering (Aggarwal, 2003; Lampe & Hauser, 2011), astronomy (Li et al., 2021), economics (Aggarwal, 2003), information engineering (Lampert, 2015), environmental studies (Krempl et al., 2019), and cinematic art studies (Sreenivasan, 2013). In particular, with the growth of data volume and the development of IT devices, relevant techniques for kernel density estimation in time-series analysis are advancing, especially in streaming data[20] analysis, where data is handled and analysed at high speed and in large volumes. The three subsections in this section provide useful techniques for solving the problems in financial time-series research using kernel density estimation are mentioned in interdisciplinary domains, not limited to the previous literature in the financial domain; visualisation of high-frequency time-evolving density distributions, prediction of future density distributions and faster density estimation without loss of accuracy.

4.1 Visualisation of high-frequency time-evolving density distributions

Detailed visualisation of the time evolution of the density distribution of data is one of the most important tasks in the various area of time-series analysis, including the financial domain. When using time series with large amounts of data or using non-stationary time series, it is not easy to comprehensively describe the time evolution of statistical features of time series from general visualisation approaches, such as functions and line plots, or time-series plots of summary statistics, such as mean and variance (Lampe & Hauser, 2011). Conversely, a comprehensive observation of the time evolution of density distributions using time-series data enables a more detailed understanding of its time-evolution pattern, which is difficult to achieve with the general approaches mentioned above (Arroyo & Maté, 2009; Krempl et al., 2019). Indeed, the effectiveness of visualising the time evolution based on kernel density estimation has already been demonstrated using large or non-stationary time series (Lampe & Hauser, 2011; Li et al., 2021).

It is also very significant from a practical and ethical point of view in data analysis. Machine learning[21] techniques, such as deep learning, are advancing rapidly in time-series analysis, leading to increasingly complex and non-linear modelling between time series and predictions based on this modelling. However, this situation is beginning to raise prac-

Figure 6: Visualisation of the time evolution of density distributions using the idea of contour plots.



Note: As in Figure 5, the fixed kernel density estimation was applied 50 times to 50 random numbers based on the standard normal distribution and visualised simultaneously. This figure, which utilises the idea of contour plots, provides a balanced view of the time evolution of the features of the density distribution from its central behaviour near the mean to the local behaviour at the tail of the distribution. The problem of too many overlapping lines, which was a problem in Figure 5 left figure, and the difficulty in observing the time evolution of the local structure in Figure 5 right figure, are resolved in a well-balanced manner.

tical and ethical issues as they increase the black box of decision-making processes, such as modelling and predictions based on increasingly complex methods (Guidotti et al., 2018). Detailed visualisation of the time evolution of a time series using density estimation leads to the discovery of its time-evolution patterns, which also can significantly contribute to explaining modelling and forecasting processes.

The points mentioned above exist in the area of interdisciplinary time-series analysis, and a visualisation method different from the previous section is also used; a visualisation method of time evolution by density estimation, which is effective for time series with a large amount of data. It is a 2D visualisation approach based on heat maps and contour plots, as shown in Figure 6, where the x-axis is represented as time, the y-axis as random variables and the colour differences and shading as a probability density. This visualisation approach makes it possible to overcome the difficulties of interpretation due to overdrawn lines in the density distribution and the difficulties of interpreting local features of density distribution, as shown in Figure 5. Several empirical examples (e.g. music and temperature data) have shown that it is effective for visualising data with non-normal distribution structures, such as multimodality and spatiotemporally large-scale data (Lampe & Hauser, 2011). Similarly, in the space science field, it has helped to empirically confirm the large time evolution and bimodality of density distributions of time series (Li et al., 2021).

Figure 6 shows the results of a test in which the idea of contour plots was applied to the same sample as in Figure 5, with the x-axis representing time, the y-axis representing the random variable and the colour differences representing the probability density. The respective time developments of the central and local structures of the density distribution, which were difficult to understand in Figure 5, can be confirmed in a balanced manner. Therefore, it is expected to be effective in visualising the time evolution of the density distribution when estimating the kernel density to high-frequency financial time series.

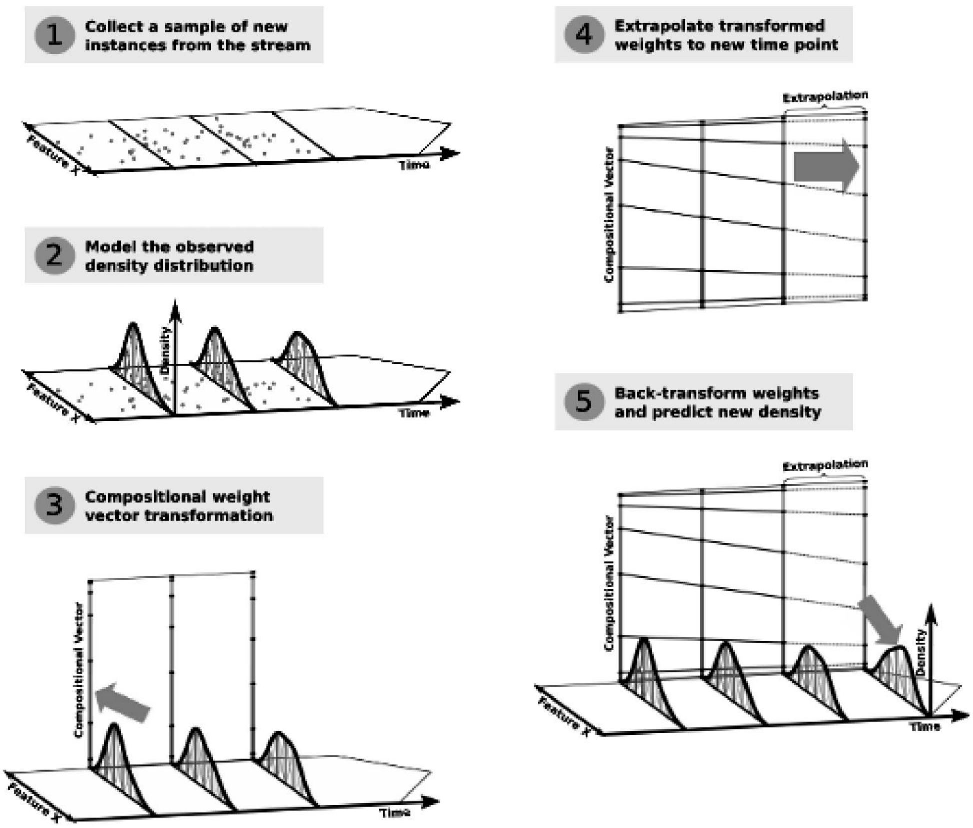
Visualising time evolution by density distributions, including the concept in Figure 6, can complement the conventional description of time evolution by summary statistics (e.g. mean and variance) without competing with the conventional description approach. Summary statistics can also be calculated simultaneously from density distributions (Bessa et al., 2012; Semeyutin & O'Neill, 2019), and their accuracy is reported to be higher than when estimated directly from the samples (He & Li, 2018). In addition, new methods to quantitatively describe time evolution based on the shape of the density distribution are available. Such approaches quantify the time evolution of the density distribution of data from the non-overlap of two distributions before and after in time (Kraemer et al., 2021; Pastore & Calcagnì, 2019), identify the distribution shape by quantifying non-normality and asymmetry (Iwamoto & Takada, 2018), and establish changes in the time-evolution pattern of the density distribution (Goswami et al., 2018). The combination of various visualisation methods, such as the above approach and popular summary statistics, enables the more detailed and comprehensive observation of the time evolution of the density distribution of time-series data.

4.2 Prediction of future density distributions

Several new challenges have been made to solve the problem of analysing the time evolution of time series by density estimation, which was pointed out in the previous section, namely the computational cost. This subsection describes one of these attempts, a technique called future density distribution prediction. This technique attempts to model the time evolution of the density distribution shape and extrapolate it one period into the future. While this does not directly reduce computational costs, it allows adaptive behaviour in advance of the future time evolution of the density distribution, thus reducing the negative impact of computational costs on the analysis and decision-making process, such as validation latency and obsolescence of analysis methods (Hofer & Krempl, 2013; Krempl et al., 2019; Lampert, 2015).

There is a common procedure for predicting future density distributions. Figure 7 shows an example of a future distribution forecast using density estimation. The pattern of the time evolution of the density distribution is modelled on a data space, such as Euclidean or Hilbert space, using modelling methods, such as regression analysis or neural networks (see Processes 2 and 3 in Figure 7). The future density distribution is extrapolated from the time weights (see Processes 4 and 5 in Figure 7). There is great diversity in what data space is used and how the pattern of the time evolution of the density distribution of a time series is modelled (Arroyo & Maté, 2009; Bessa et al., 2012; He & Li, 2018; Krempl et al., 2019; Lampert, 2015), and there is currently no de facto standard-like procedure. Demonstrations using time-series data from various domains have shown the effectiveness of predicting their future density distribution, video and image data (Lampert, 2015), credit score data (Krempl et al., 2019), environmental health data (Krempl et al., 2019), wind data (Bessa et al., 2012; He & Li, 2018), and exchange rate data (Arroyo & Maté, 2009). Therefore, as the literature is not mentioned individually in

Figure 7: An example of a probability density-based approach to predicting future distribution shapes



Note: This figure was adapted from Figure 1 in Krempl et al. (2019). Proess 1 illustrates the observed time-series data X. Proess 2 illustrates the observed density distribution modelled as an expansion of the density function. Proess 3 illustrates the compositional vectors formed by the temporal weights of its basis functions. Proess 4 illustrates the time-weighted compositional vectors modelled and extrapolated into the future. Proess 5 illustrates the extrapolated compositional vectors inverted to predict the new density distribution.

this paper, see them for details on the methodology.

Currently, future density distribution forecasting techniques are novel and embryonic. Each proposed method has good and bad points, and its unthinking application to financial time series is controversial. Some examples show the effectiveness of future density distribution forecasting techniques only for time-series data with monotonically time-evolving density distribution patterns (Krempl et al., 2019; Lampert, 2015) and only for time-series data with a repeating context (Arroyo & Maté, 2009). With current techniques, it is expected to be difficult to capture rare and extreme temporal evolution of density distributions. Therefore, its application to highly non-stationary and evolving data, such as financial time series, is a challenging area that is expected to develop in future density distribution prediction techniques.

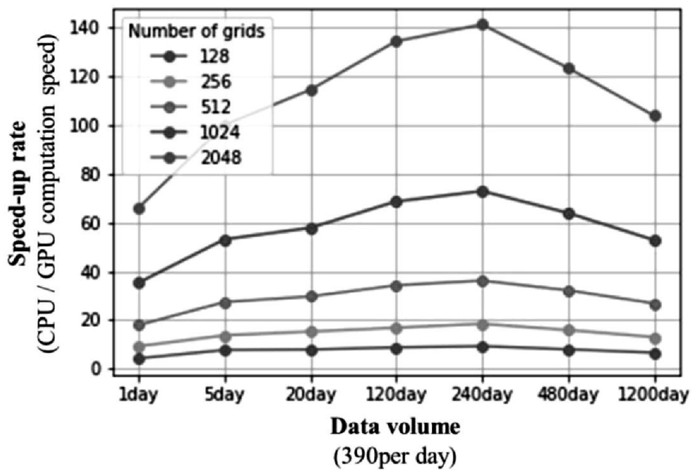
4.3 Faster density estimation without loss of accuracy

A direct solution to the computational costs of continuous density estimation in time-series analysis has also been suggested. This solution uses distributed parallel processing[22] technology, such as GPUs[23], to speed up algorithms. Distributed parallel processing technology enables speed-up without compromising the accuracy and detail of the time evolution of the density distribution, such as simplifying the calculation results by approximation and reducing the number of random variable positions in the density distribution to be estimated. Indeed, the use of GPUs has enabled real-time visualisation of the time evolution of the density distribution of streaming data (Daae Lampe & Hauser, 2011; Lampe & Hauser, 2011).

The effect of the improved computational speed of the density estimation method is examined. Figure 8 shows how much faster the GPU-based fixed kernel density estimation method is compared to the CPU-based fixed kernel density estimation method. How the results change with the number of random variable grids required and the number of data is also investigated. As a result, a GPU speed-up effect was observed in all cases. The speed-up effect increases for cases with a larger number of grids of random variables, that is, those analysing more detailed density distribution shapes. Therefore, speeding up the process of analysing the time evolution of the density distribution on a GPU has proved very beneficial.

However, applying distributed parallel processing technologies, such as GPUs, to time-series analysis, including in the financial domain, will reduce this benefit. The key to GPU acceleration lies in managing GPU memory and data transmission and reception between CPU and GPU (Michailidis & Margaritis, 2013). Density estimation algorithms have been improved to suit distributed parallel processing techniques according to the analysis method, such as minimising GPU memory usage and GPU-CPU communication (Lampe & Hauser, 2011). No previous studies have implemented density estimation methods, such as adaptive kernel density estimation mentioned in Section 2, which are robust to anomalies and have high tail estimation accuracy on GPUs and used them for analysing the time evolution of time series. The active introduction of distributed parallel computing technology, such as GPUs, would greatly contribute to advancing research

Figure 8: Effects of GPUs on speeding up density estimation.



Note: The figure shows how many times faster the univariate GPU-based fixed kernel density estimation (y-axis) was from the univariate CPU-based fixed kernel density estimation. Y-axis results are the median of 50 simulation results. x-axis shows the amount of random data obtained from the standard normal distribution. A high-frequency financial time series (data at 1-minute intervals) of the New York Stock Exchange in the USA, the largest in the world, was assumed and a volume of 390 data per day was used (for example, 1 day means 390 data were used, 5 days means 1950 data were used). The legend (different colours) indicates how many grids of random variables for which the probability density was estimated. CPU (Intel Xeon) and GPU (NVIDIA Tesla P100 PCIe) were used through a python package (PyCuda) on Google Colaboratory. For the setup of GPU-based kernel density estimation, 256 threads per block and the most naive algorithm (Michailidis & Margaritis, 2013) were used. As a result, a GPU speed-up effect was observed in all cases. The speed-up effect increases for cases with a larger number of grids of random variables, that is, those analysing more detailed density distribution shapes.

on analysing the time evolution of density distributions using financial time series.

5 Conclusion

This study's approach was to analyse the comprehensive shape of the density distribution of returns directly from the sample and the time evolution of the shape, without making arbitrary assumptions and without summarising the information. The previous literature was reviewed to explore the procedural features and problems of analysing the density distribution of asset price returns. This study is unique in that it considered applying non-parametric probability density estimation methods to high-frequency financial data in light of recent trends in the financial domain. Furthermore, an interdisciplinary survey of data analysis techniques that could address the current challenges identified in the review mentioned above was conducted, and the effectiveness of these techniques was tested.

The results show common procedural features of density distribution analysis of asset price returns in the financial domain. However, there is disagreement on the bandwidth selection method, and the choice should suit the research objectives. The study also identified some remaining issues when using non-parametric probability density estimation in the area of analysis of financial data, which is becoming more frequent: the existence of outliers, high computational costs, and the difficulty of visualisation. Next, useful techniques for solving problems in financial time-series research using kernel density estimation were mentioned in three categories: visualisation of high-frequency time-evolving density distributions, prediction of future density distributions, and speed-up of density estimation without loss of accuracy. This study showed that algorithms using distributed parallel processing techniques, such as GPUs, can speed up the time evolution of density distributions without compromising their accuracy or detail, and in addition, without simplifying the calculation results through approximation or reducing the number of random variable positions in the density distribution to be estimated.

This study provides a better understanding of density distribution analysis of returns using non-parametric density estimation and supports the construction of an analytical design consistent with the research objectives. In addition, it removes barriers to introducing non-parametric density estimation methods in studies of higher frequency financial data in general, including research on price asset returns. Empirically investigating the statistical properties of asset price-return dynamics could help develop models and theories of future asset price dynamics, thereby contributing to rapid risk control by market participants and overall market stability.

As confirmed in section 3, empirical analysis of the time evolution of density distributions in financial domain research is currently not very abundant. Further empirical analysis of various research hypotheses using financial time series data, especially high-frequency data, for various markets and timescales is required to clarify such analysis's effectiveness in financial domain research. Applying the useful analytical techniques identified in this paper to various financial time series data makes it possible to discover characteristic patterns in the time evolution of density distributions in financial time series that have not been visible before. This work will consequently contribute to good modelling of their time evolution patterns for predicting future density distributions discussed in Section 4 and refining research designs suited to the financial domain. Empirical analysis of the time evolution of density distributions using real data from high-frequency financial transactions and the unique and novel fact-finding derived from this analysis should be the subject of my future research.

Acknowledgements

This work was supported by JST SPRING program, Grant Number JPMJSP2139.

Note

1) Financial econometrics is a discipline that applies data analysis methods to financial time series to understand the nature of financial time series empirically and to predict the future of financial time series.

2) High-frequency financial data refers to time-series data recorded at small time-scale intervals, such as per transaction, second, or minute.

3) Returns represent the rate of change in asset prices in financial markets, and in many cases, the logarithmic rate of change is used due to its analytical validity.

4) Moment statistics are a quantitative measure of probability distribution shape. The first moment represents the expected value, the second moment the variance, the third moment the skewness and the fourth moment the kurtosis.

5) Non-stationary data refers to time-series data whose statistical properties evolve with time, and such time series are highly difficult to predict (model).

6) Non-parametric approaches differ from parametric approaches and refer to statistical analysis approaches that are not based on arbitrary parameter assumptions of probability distributions.

7) A true density function is a function that represents the relative distribution of frequencies that a continuous random variable would essentially follow. The integral value of an interval of the function is the probability that the random variable corresponding to that interval will occur.

8) Multimodality refers to the property of a probability distribution to have multiple peaks.

9) Fat-tails refer to the property of a probability distribution to have a thick tail part of it, in other words, to be relatively prone to rare phenomena.

10) A kernel function indicates a function that applies a fixed core mathematical process to input data.

11) Outliers refer to data located far from the central position of the observed data. In some cases, outliers are due to intrinsic properties of the data, while in others, they are data errors (noise) that distort the results.

12) The pilot estimator refers to the preliminary estimator required in the calculation algorithm's pilot (intermediate) stage.

13) The normal or Gaussian distribution refers to a symmetrical, continuous probability density distribution that takes a bell-like shape.

14) Derivatives refer to financial contracts or instruments that are linked to the performance of an underlying financial asset, such as a financial index or benchmark.

15) The sliding window method is one of the window models used in data stream analysis, where data arrives in huge or theoretically infinite numbers, and refers to dividing data into a fixed or adaptive number of windows in order of arrival and analysing each window separately. Other typical models include landmark window, damped window and sloping window models (Mansalis et al., 2018; Silva et al., 2013).

16) R, Python and MATLAB are some of the high-level languages, in which statistical analysis can be carried out with relative simplicity.

17) Intraday data refers to high-frequency data with time intervals smaller than daily in financial time-series data, such as seconds and minutes.

18) Algorithmic trading refers to automated trading carried out mechanically according to some rules, while high-frequency trading refers to high-speed trading completed in a short period, such as a few seconds.

19) Data pre-processing refers to processing, deleting or organising raw data prior to analysis to meet the research objectives. Excessive pre-processing can distort the nature of the data, and intermediate processing to suit the analysis design is desirable.

20) Streaming data is continuously generated time series, which is large and high-frequency data that is theoretically infinite.

21) Machine learning is the general term for disciplines and tools that perform analytical tasks, such as classification and prediction of events with data using algorithms using statistical methods.

22) Distributed parallel processing refers to the distributed execution of tasks using a large number of computational cores. When processing a large number of tasks using computing cores of the same performance, distributed execution of tasks tends to be faster than serial execution of tasks on a single computing core.

23) GPU stands for graphics processing unit and originally refers to a specialised processor designed to make a device's screen display faster and smoother. In recent years, GPUs have also been used for data analysis, such as machine learning, as they can simultaneously process large numbers of data.

References

Abramson, I. S. (1982). On bandwidth variation in kernel estimates—A square root law. *Annals of Statistics, 10*(4), 1217-1223.

Aggarwal, C. C. (2003). A framework for diagnosing changes in evolving data streams. *Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data*, 575-586.

Amaya, D., Christoffersen, P., Jacobs, K., & Vasquez, A. (2015). Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics, 118*(1), 135-167.

Arnerić, J. (2020). Realized density estimation using intraday prices. *Croatian Review of Economic, Business and Social Statistics, 6*(1), 1-9.

Arroyo, J., & Maté, C. (2009). Forecasting histogram time series with k-nearest neighbours methods. *International Journal of Forecasting, 25*(1), 192-207.

Bagnato, L., De Capitani, L., & Punzo, A. (2014). Testing serial independence via density-based measures of divergence. *Methodology and Computing in Applied Probability, 16*(3), 627-641.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2009). Realised kernels in practice: Trades and quotes. *The Econometrics Journal, 12*, 1-33.

Bessa, R. J., Miranda, V., Botterud, A., Wang, J., & Constantinescu, E. M. (2012). Time adaptive conditional kernel density estimation for wind power forecasting. *IEEE Transactions on Sustainable Energy, 3*(4), 660-669.

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2), 353-360.

Breiman, L., Meisel, W., & Purcell, E. (1977). Variable kernel estimates of multivariate densities. Technometrics: *A Journal of Statistics for the Physical, Chemical, and Engineering Sciences, 19* (2), 135-144.

Brownlees, C. T., & Gallo, G. M. (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics & Data Analysis, 51*(4), 2232-2245.

Cai, C. X., Kim, M., Shin, Y., & Zhang, Q. (2018). FARVaR: Functional autoregressive value-at-risk. *Journal of Financial Economics, 17*(2), 284-337.

Campbell, J. Y., Lo, A. W., Craig MacKinlay, A., & Whitelaw, R. F. (1998). The econometrics of financial markets. *Macroeconomic Dynamics, 2*(4), 559-562.

Chen, J., Wang, Y., & Ren, X. (2022). Asymmetric effects of non-ferrous metal price shocks on clean energy stocks: Evidence from a quantile-on-quantile method. *Resources Policy, 78.* https://doi.org/10.1016/j.resourpol.2022.102796

Chen, Y.-C. (2017). A tutorial on kernel density estimation and recent advances. *Biostatistics & Epidemiology, 1*(1), 161-187.

Čížek, P., & Sadıkoğlu, S. (2020). Robust nonparametric regression: A review. *Wiley Interdisciplinary Reviews: Computational Statistics, 12*(3), e1492.

Efromovich, S. (2010). Orthogonal series density estimation. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(4), 467-476.

Eilers, P. H. C., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science, 11*(2), 89-121.

Engle, R. F. (2000). The econometrics of ultra-high-frequency data. *Econometrica, 68*(1), 1-22.

Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. CRC Press.

Falkenberry, T. N. (2002). High frequency data filtering. Working Paper, Tick Data.

Faraway, J. J., & Jhun, M. (1990). Bootstrap choice of bandwidth for density estimation. *Journal of the American Statistical Association, 85*(412), 1119-1122.

Friedman, J. H., & Tukey, J. W. (1974). A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers. Institute of Electrical and Electronics Engineers, C-23*(9), 881-890.

Fukunaga, K. (2013). *Introduction to statistical pattern recognition*. Elsevier.

Goswami, B., Boers, N., Rheinwalt, A., Marwan, N., Heitzig, J., Breitenbach, S. F. M., & Kurths, J. (2018). Abrupt transitions in time series with uncertainties. *Nature Communications, 9*(1), 48.

Gu, C., Kurov, A., & Wolfe, M. H. (2018). Relief rallies after FOMC announcements as a resolution of uncertainty. *Journal of Empirical Finance, 49*, 1-18.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys, 51*(5), 1-42.

Gurrib, I., Elshareif, E. E., & Kamalov, F. (2020). The effect of energy cryptos on efficient portfolios of key energy listed companies in the S&P Composite 1500 Energy Index. *International Journal of Energy, 10*(2), 1-15.

He, Y., & Li, H. (2018). Probability density forecasting of wind power using quantile regression neural network and kernel density estimation. *Energy Conversion & Management, 164*, 374-384.

Hofer, V., & Krempl, G. (2013). Drift mining in data: A framework for addressing drift in classification. *Computational Statistics & Data Analysis, 57*(1), 377-391.

Iwamoto, N., & Takada, T. (2018). Nonparametric estimation of the size premium using high-frequency data. *The Business Review, 68*(4), 111-126.

Izenman, A. J. (1991). Review papers: Recent developments in nonparametric density estimation. *Journal of the American Statistical Association, 86*(413), 205-224.

Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(1), 4-37.

Janssen, P., Marron, J. S., Veraverbeke, N., & Sarle, W. (1995). Scale measures for bandwidth selection. *Journal of Nonparametric Statistics, 5*(4), 359-380.

Jones, M. C., Marron, J. S., & Sheather, S. J. (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association, 91*(433), 401-407.

Kraemer, B. M., Pilla, R. M., Woolway, R. I., Anneville, O., Ban, S., Colom-Montero, W., Devlin, S. P., Dokulil, M. T., Gaiser, E. E., Hambright, K. D., Hessen, D. O., Higgins, S. N., Jöhnk, K. D., Keller, W., Knoll, L. B., Leavitt, P. R., Lepori, F., Luger, M. S., Maberly, S. C., … Adrian, R. (2021). Climate change drives widespread shifts in lake thermal habitat. *Nature Climate Change, 11*(6), 521-529.

Krempl, G., Lang, D., & Hofer, V. (2019). Temporal density extrapolation using a dynamic basis approach. *Data Mining and Knowledge Discovery, 33*(5), 1323-1356.

Lampe, O. D., & Hauser, H. (2011a). Curve density estimates. *Computer Graphics Forum, 30*(3), 633-642.

Lampe, O. D., & Hauser, H. (2011b). Interactive visualization of streaming data with kernel density estimation. *2011 IEEE Pacific Visualization Symposium*, 171-178.

Lampert, C. H. (2015). Predicting the future behavior of a time-varying probability distribution. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 942-950.

Li, D., Wang, P., Zhu, W. W., Zhang, B., Zhang, X. X., Duan, R., Zhang, Y. K., Feng, Y., Tang, N. Y., Chatterjee, S., Cordes, J. M., Cruces, M., Dai, S., Gajjar, V., Hobbs, G., Jin, C., Kramer, M., Lorimer, D. R., Miao, C. C., … Zhu, Y. (2021). A bimodal burst energy distribution of a repeating fast radio burst source. *Nature, 598*(7880), 267-271.

Liu, B., Zhou, C., & Zhang, X. (2019). A tail adaptive approach for change point detection. *Journal of Multivariate Analysis, 169*, 33-48.

Liu, Q., Xu, J., Jiang, R., & Wong, W. H. (2021). Density estimation using deep generative neural networks. *Proceedings of the National Academy of Sciences of the United States of America, 118* (15), e2101344118.

Mansalis, S., Ntoutsi, E., Pelekis, N., & Theodoridis, Y. (2018). An evaluation of data stream clustering algorithms. *Statistical Analysis and Data Mining, 11*(4), 167-187.

Marron, J. S., & Nolan, D. (1988). Canonical kernels for density estimation. *Statistics & Probability Letters, 7*(3), 195-199.

Mei, D., Liu, J., Ma, F., & Chen, W. (2017). Forecasting stock market volatility: Do realized skewness and kurtosis help? *Physica A: Statistical Mechanics and Its Applications, 481*, 153-159.

Michailidis, P. D., & Margaritis, K. G. (2013). Accelerating kernel density estimation on the GPU using the CUDA framework. *Applied Mathematical Sciences, 7*, 1447-1476.

Müller, H.-G. (2012). *Nonparametric regression analysis of longitudinal data*. Springer Science & Business Media.

Müller, U. (2001). The Olsen filter for data in finance. Working Paper, Olsen and Associates.

Pastore, M., & Calcagnì, A. (2019). Measuring distribution similarities between samples: A distribution-free overlapping index. *Frontiers in Psychology, 10*, 1089.

Prakasa Rao, B. L. (2014). *Nonparametric functional estimation*. Academic Press.

Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Annals of Statistics, 3*(1), 1-14.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. Scandinavian Journal of Statistics, *Theory and Applications, 9*(2), 65-78.

Scott, D. W. (1979). On optimal and data-based histograms. *Biometrika, 66*(3), 605-610.

Scott, D. W. (1985). Frequency polygons: Theory and application. *Journal of the American Statistical Association, 80*(390), 348-354.

Scott, D. W. (2015). *Multivariate density estimation: Theory, practice, and visualization*. John Wiley & Sons.

Scott, D. W., & Sain, S. R. (2005). Multidimensional density estimation. In C. R. Rao, E. J. Wegman, & J. L. Solka (Eds.), *Handbook of Statistics* (Vol. 24, pp. 229-261). Elsevier.

Scott, D. W., & Terrell, G. R. (1987). Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association, 82*(400), 1131-1146.

Semeyutin, A., & O'Neill, R. (2019). A brief survey on the choice of parameters for: "Kernel density estimation for time series data." *The North American Journal of Economics and Finance, 50*, 101038.

Sheather, S. J. (2004). Density estimation. *Statistical Science, 19*(4), 588-597.

Sheather, S. J., & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, 53*(3), 683-690.

Silva, J. A., Faria, E. R., Barros, R. C., Hruschka, E. R., de Carvalho, A. C. P. L. F., & Gama, J. (2013). Data stream clustering: A survey. *ACM Computing Surveys, 46*(1), 1-31.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall.

Sindagi, V. A., & Patel, V. M. (2018). A survey of recent advances in CNN-based single image crowd counting and density estimation. *Pattern Recognition Letters, 107*, 3-16.

Sreenivasan, S. (2013). Quantitative analysis of the evolution of novelty in cinema through crowdsourced keywords. *Scientific Reports, 3*, 2758.

Takada, T. (2008). Asymptotic and qualitative performance of non - parametric density estimators: A comparative study. *The Econometrics Journal, 11*(3), 573-592.

Takada, T. (2009). Extracting phases of financial markets. *International Workshop on Data-Mining and Statistical Science, DMSM-A901*, 27-30.

Tay, A. (2015). A brief survey of density forecasting in macroeconomics. *Macroeconomic Review*, October.

Taylor, C. C. (1989). Bootstrap choice of the smoothing parameter in kernel density estimation. *Biometrika, 76*(4), 705-712.

Tsay, R. S. (2016). Some methods for analyzing big dependent data. *Journal of Business & Economic Statistics, 34*(4), 673-688.

Verousis, T., & Gwilym, O. A. (2010). An improved algorithm for cleaning ultra high-frequency data. *Journal of Derivatives & Hedge Funds, 15*(4), 323-340.

Wang, R., Dearing, J. A., Langdon, P. G., Zhang, E., Yang, X., Dakos, V., & Scheffer, M. (2012). Flickering gives early warning signals of a critical transition to a eutrophic lake state. *Nature, 492* (7429), 419-422.

Wang, X., Tsokos, C. P., & Saghafi, A. (2018). Improved parameter estimation of time dependent kernel density by using artificial neural networks. *The Journal of Finance and Data Science, 4*(3), 172-182.

Wang, Z., & Scott, D. W. (2019). Nonparametric density estimation for high - dimensional data—Algorithms and applications. *Wiley Interdisciplinary Reviews: Computational Statistics, 11*(4), e1461.

Wasserman, L. (2006). *All of nonparametric statistics*. Springer Science & Business Media.

Woodroofe, M. (1970). On choosing a delta-sequence. *Annals of Mathematical Statistics, 41*(5), 1665-1671.