The Pennsylvania State University
The Graduate School

DENSITY ESTIMATION FOR SOME SEMIPARAMETRIC MODELS

A Dissertation in
Statistics
by
Manuel Dario Hernandez Bejarano

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2024

The dissertation of Manuel Dario Hernandez Bejarano was reviewed and approved by the following:

Zhibiao Zhao
Associate Professor of Statistics
Dissertation Advisor
Chair of Committee

Runze Li
Professor of Statistics

Lingzhou Xue
Professor of Statistics

Andres Aradillas-Lopez
Associate Professor of Economics

Bing Li
Director of Graduate Studies

Abstract

The classical nonparametric kernel density estimator has been widely used to estimate the marginal density of a variable of interest in fields such as finance and economics. However, it has several limitations, including a slow convergence rate, which becomes particularly problematic in small sample sizes. This dissertation document is concerned with studying more efficient density estimations for the marginal density of two important semiparametric models. We show that the proposed estimators exhibit appealing properties that are absent in the classical estimator.

In the first project, chapter 2, motivated by the slow convergence rate of classical nonparametric kernel density estimator, we study more efficient density estimation and density derivative estimation for the marginal density of nonparametric regression models. We show that in the presence of unknown nonparametric regression function, the proposed density and density derivative estimators can achieve parametric convergence rate, $\sqrt{n}$, and possess several appealing properties which the classical estimator lacks. Also, in the absence of nonparametric regression function, when the noise is normally distributed, the proposed method performs as well as if we have known the model and estimated the density using the maximum likelihood method. Based on the proposed density estimator, we further propose a more powerful density-based specification test for the nonparametric regression function. Our extensive numerical studies show that the proposed density estimator, density derivative estimator, and specification test significantly outperform existing ones.

In the second project, chapter 3, we study a more efficient density estimation for the stationary density of nonparametric autoregressive conditional heteroscedasticity (NARCH) models. These models are important tools in analyzing time series, specifically in economic and financial applications where the goal is modeling and understanding the volatility of the statistical data since this volatility appears to change over time and exhibit clustering[1]. We demonstrate that in the presence of an unknown nonparametric variance structure, we can establish the $\sqrt{n}$ consistency of the proposed density estimator, improving this way the widely used nonparametric kernel density estimator whose rate of convergence is inferior. A numerical study confirms the results. The density estimator is applied to the S&P 500 Index data. Finally, we showcase a practical implementation

[1]In this context, the term clustering can be defined as periods of swings interspersed with periods of relative calm [Bollerslev et al. (1994)].

of the proposed density estimator in quantile regression. Specifically, we propose to get a more accurate estimate of the limiting variance of the estimated coefficients in a quantile regression model whose errors follow a nonparametric autoregressive conditional heteroscedastic structure. We perform a simulation study, which shows that using the new density estimator leads to a more accurate estimation of this asymptotic variance compared to the results obtained using the classical density estimator. To illustrate the application of this methodology in estimating the asymptotic variance, we apply it to the monthly inflation rate of the United States.

Finally, Chapter 4 summarizes the main conclusions of the projects outlined in this document, as well as two potential avenues for future research in density estimation in the context of time series.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my sincere gratitude towards my Ph.D. advisor, Dr. Zhibiao Zhao, for his guidance and support throughout these past years. This dissertation could not have been written without his expertise as a great researcher and teacher. I am thankful for his help, patience, and enthusiasm throughout this process. Additionally, I would like to thank Dr. Runze Li, Dr. Lingzhou Xue, and Dr. Andres Aradillas-Lopez for being my committee members and providing helpful, insightful comments and suggestions. I would also like to extend my thanks to all the Department of Statistics faculty at Pennsylvania State University for their exceptional lectures and guidance during my studies. Furthermore, I would also like to express my appreciation to my best friend, Deicy Cristiano, who has supported and helped my family and me since our arrival at the university. Finally, I would like to express a special thanks to my wife, Johana Molano, for her unwavering support and care throughout these years, for which my mere expression of gratitude is not enough.

Chapter 1 Introduction

1.1 What is density estimation?

The marginal density function plays an important role in statistical analysis as it contains all distributional information about the data. Silverman (1986) indicates that "*A very natural use of density estimates is in the informal investigation of the properties of given set of data. Density estimates can give valuable indication of some features as skewness and multimodality in the data. In some cases they will yield conclusion that may then be regarded as self-evidently true, while in others all they will do is to point the way to further analysis and/or data collection*".[1]

One way to estimate the marginal density function is by using a parametric approach, which assumes that the data Y has been taken from a known parametric family of distributions, for example, the normal distribution with parameters μ and σ^2. In this case, the density function f_Y underlying the data could be estimated by finding estimates of the mean and variance from the data using appropriate methods, such as the maximum likelihood method, and substituting these estimates into the expression of the normal density function, see Silverman (1986) for more details. However, as noted by Li and Racine (2007), the weakness of this parametric approach is that, before the estimation process, the researcher must specify the exact parametric functional form for the density that needs to be estimated. In other words, the parametric approach is somewhat circular as we initially set out to estimate an unknown density but must first assume that the density is, in fact, known. Moreover, having based our estimate on the assumption that the density is a member of a known parametric family of distributions, we need to consider the possibility that our parametric approach can be misspecified, that is to say,

[1] See Silverman (1986) for more discussion.

not consistent with the population from which the data came.

Once we estimate the density by the parametric approach, it can be convenient always to test whether the underlying distributional assumption is adequate. However, if we reject the distributional assumption of the data, we will not have any clear alternative [Li and Racine (2007)]. For instance, we can reject the normality assumption, but we will still be left with many candidate distributions to choose from. In such cases, embracing the nonparametric approach emerges as a viable recourse for the researchers, circumventing the issues of the parametric approach described above.

The other way to estimate the marginal density function is by using nonparametric approaches. These approaches are used to avoid problems that arise from specifying the parametric distribution of the observed data before estimation. Instead of assuming that we know the exact functional form of the density that requires to be estimated, we assume that it satisfies certain regularity conditions such as smoothness and differentiability. However, this approach requires more available data to achieve the same degree of precision as a correctly specified parametric model because it imposes less structure on the functional form of the density function than parametric approaches, see Li and Racine (2007) or Racine (2008) for more details. One of the most commonly used nonparametric approaches is the traditional nonparametric kernel density estimator, the so-called Parzen–Rosenblatt [Silverman (1986)]. Given $Y_1, \ldots, Y_n$, we have

$$\tilde{f}_Y(y) = \frac{1}{nb_n} \sum_{i=1}^{n} K\left(\frac{y - Y_i}{b_n}\right), \tag{1.1}$$

where b_n is the bandwidth, also called the smoothing parameter, and $K(\cdot)$ represents a kernel function. Essentially, $\tilde{f}_Y(y)$ smoothes each data point Y_i, $1 \leq i \leq n$, into small density bumps and then sums all these small bumps together to obtain the final density estimate. This is the estimator expressed in (2.2) in Chapter 2 or in equation (3.2) in Chapter 3. There is extensive literature on this estimator, depending on the structure of the data (see Chapters 2 and 3). Based on Silverman (1986) and Li and Racine (2007), we have the following observations on this estimator:

- Using Taylor expansion, we have that the bias of $\tilde{f}_Y(y)$ is

$$\text{Bias}\left(\tilde{f}_Y(y)\right) = \frac{b_n^2 f_Y''(y)}{2} \int_{\mathbb{R}} u^2 K(u) du + o(b_n^2).$$

This means that when $b_n \to 0$, the bias is shrinking at a rate $O\left(b_n^2\right)$ [see Li and Racine (2007)]. Also, by similar argument, the variance of $\tilde{f}_Y(y)$ is

$$\operatorname{Var}\left(\tilde{f}_Y(y)\right) = \frac{1}{nb_n} f_Y(y) \int_{\mathbb{R}} K^2(u)du + o\left(\frac{1}{nb_n}\right).$$

Thus, the variance shrinks at rate $O\left(1/(nb_n)\right)$ when we allow $nb_n \to \infty$ as $n \to \infty$. According to Racine (2008), the term nb_n is sometimes referred to as the *effective sample size*. Also, the condition that $nb_n \to \infty$ as $n \to \infty$ simply requires that as we get more information ($n \to \infty$) we average over a narrower region ($b_n \to 0$) but the amount of *local information* (nb_n) must increase simultaneously. Finally, it is worth noting that at the point where the density value is large, the variance of $\tilde{f}_Y(y)$ is also large.

The above two formulas for bias and variance of $\tilde{f}_Y(y)$ have the pointwise property, which means they hold at any point y.

- Under appropriate conditions [see Li and Racine (2007)],

$$\sqrt{nb_n}\left[\tilde{f}_Y(y) - f_Y(y) - \frac{b_n^2 f_Y''(y)}{2}\int_{\mathbb{R}} u^2 K(u)du\right] \Rightarrow N\left(0, f_Y(y)\int_{\mathbb{R}} K^2(u)du\right).$$

That is, the estimator $\tilde{f}_Z(z)$ is $\sqrt{nb_n}$-consistent, and the slow convergence rate can be problematic for small sample sizes n.

- $\tilde{f}_Y(y)$ has some advantages and disadvantages [see Silverman (1986)]. On the one hand, $\tilde{f}_Y(y)$ does not depend on the underlying model of Y_i, $1 \leq i \leq n$; thus, it is robust against some model assumptions. However, on the other hand, it does not take into account the specific structure of the data Y_i, so it may not be the most efficient estimator available.

- The estimator $\tilde{f}_Y(y)$ can have a problem when used with data from long-tailed distributions, such as those discussed in Chapter 3. Since b_n is constant for all observations in the sample, it results in a tendency for spurious noise that appears in the tails of the estimates. If the estimates are smoothed adequately to address this issue, then sometimes important details in the central part of the distribution can be masked. To address this problem, various adaptive methods have been proposed, such as the nearest neighbor method or the variable kernel method; see

Silverman (1986) for more details. In this document, we will not consider these kinds of adaptive methods; we only focus on the estimator $\tilde{f}_Y(y)$.

1.2 Motivation of the research

Because of the fundamental role of the marginal density function, the construction of an estimate of this density function from the observed data is an important aspect of applied data analysis in various fields. One of the most basic tests in applied data analysis is to test the equality of two distributions or moments thereof [see Li and Racine (2007)]. For instance, in economics, we study inflation rate distributions and how they differ across regions, expenditure groups, and over time [Silverman (1986); Liao and Stachurski (2015)]. Another example can be found in financial econometrics, where the goal can be to investigate and compare the distributions of returns from financial markets to obtain valuable information about the underlying processes driving these financial markets [e.g., Zhao (2008)]. Although the density function is often the primary object of interest, its estimation also plays an essential role in creating other objects, such as certain special integral functionals of the marginal density [e.g., Kerkyacharian and Picard (1996)], density-based specification test or model validation problems [e.g., Aït-Sahalia (1996)], or estimation for the asymptotic variance of the estimated coefficients in a quantile regression model [Koenker (2005)], among others.

As mentioned in Section 1.1, the traditional nonparametric kernel density estimator is one of the most widely used density estimators. This estimator does not depend on the underlying model of the stochastic process of interest, and thus, it is robust against some model assumptions. However, since the nonparametric kernel density estimator does not consider the specific structure of the process, it has a slow convergence rate of $\sqrt{nb_n}$. This slow convergence rate can be problematic for small sample sizes, as observed in the simulation studies in Chapters 2 and 3.

When we additionally consider the structure for the stochastic process of interest, density estimation can be used more cleverly to obtain better rates of convergence, in particular $\sqrt{n}$-consistency. This approach has been proposed for some parametric models, including moving average process [Saavedra and Cao (1999, 2000); Schick and Wefelmeyer (2004)], nonlinear autoregressive models of order one with constant variance [Kim and Wu (2007)], invertible linear process [Schick and Wefelmeyer (2007)], nonlinear

autoregressive conditional heteroscedastic model with parametric noise [Zhao (2010)], nonlinear autoregressive model with nonparametric noise [Kim et al. (2015)], and, more recently, autoregressive time series models for which the conditional mean and variance have a parametric specification [Truquet (2019)], among others. All of these approaches rely on the parametric form of the model. However, such parametric assumptions can hardly be justified in many applications, and there could be a risk of mis-specification.

As mentioned in the last paragraph, to get $\sqrt{n}$-consistent density estimation, the authors make use of parametric models for the structure of the stochastic process, which requires specifying functional forms for objects being estimated. On the other hand, nonparametric models do not require this. This approach is known as *nonparametric regression* or *nonparametric smoothing* in the regression framework [see Li and Racine (2007)]. An appealing property of using nonparametric models for the underlying model of the process is that they relax the parametric assumptions imposed on the data-generating process and instead allow the data to determine the adequate model.

In the past few decades, nonparametric and semiparametric models for the structure of the stochastic process have attracted a tremendous amount of attention in many areas of theoretical and applied statistics, as evidenced by the extensive literature existent, see Racine (2008) and references therein. As pointed out by Racine (2008), semiparametric methods make up some of the more popular methods for flexible estimation. Semiparametric models are formed by combining parametric and nonparametric models in a specific manner. These types of models are applicable in settings where totally nonparametric models may not perform well, for example, when the curse of dimensionality has led to highly variable estimates or when the researcher wishes a nonparametric form for the regressor and a parametric form for the structure of the error term in the model. Therefore, we can consider semiparametric models as a compromise between fully nonparametric and fully parametric specifications. A diversity of semiparametric models have been studied. In this dissertation document, considering two semiparametric models for the structure of the stochastic process of interest, we propose to develop $\sqrt{n}$-consistent density estimates. Now, we describe the two models. The first model, discussed in Chapter 2, is considered semiparametric in the sense that the regression function is nonparametric while the error term distribution is parametric. Similarly, the second model, presented in Chapter 3, is also semiparametric as it assumes a nonparametric structure for the autoregressive conditional variance and a parametric error term distribution.

1.3 Contribution of the Thesis

The main focus of this dissertation is density estimation. We aim to develop $\sqrt{n}$-consistent density estimates for the marginal density of two important semiparametric models, as these models offer flexible estimation. We show that these proposed density estimators have several appealing properties, which the widely used nonparametric kernel density estimator lacks, including a better convergence rate.

In Chapter 2, we present a more efficient density estimation and density derivative estimation for the marginal density of nonparametric regression models, assuming that the noise has a parametric distribution. Based on the new density estimator, we further propose a more powerful density-based specification test for the nonparametric regression function.

In Chapter 3, we study a more efficient density estimation for the stationary density of nonparametric autoregressive conditional heteroscedasticity (NARCH) models, assuming that the noise follows a standard normal distribution. Based on the proposed density estimator, we further propose to use it to get a more accurate estimate of the limiting variance of the estimated coefficients in a quantile regression model, whose errors follow the nonparametric ARCH structure. We achieve this by using the proposed efficient estimation of the marginal density to estimate the density function of the error terms $f_Z(z)$, of the quantile regression model.

Chapter 2
Density Estimation and Specification Testing for Nonparametric Regression Models

2.1 Introduction

Consider observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ from the nonparametric regression model

$$Y = \mu(X) + e \quad \text{with} \quad \mathbb{E}(e) = 0, \tag{2.1}$$

where $\mu(\cdot)$ is an unknown function, and e is independent of X. While linear models or more generally nonlinear parametric models enjoy better interpretability, it is often difficult to known in advance such parametric forms and there could always be a potential risk of mis-specification using, for example, a linear model to describe a highly nonlinear phenomenon. In contrast to parametric models, the nonparametric model (2.1) can let the data speak for themselves by imposing no specific structure on $\mu(\cdot)$. Thanks to this appealing feature, the nonparametric regression model (2.1) has extensive applications in economics, finance, and statistics [Fan and Gijbels (1996); Li and Racine (2007)].

As mentioned in Section 1.1, the density function plays an important role in statistical analysis as it contains all distributional information about the data. Denote by $f_Y(y)$ the density of Y. Given $Y_1, \ldots, Y_n$, the classical nonparametric kernel density estimator [see Silverman (1986)] of $f_Y(y)$ is

$$\tilde{f}_Y(y) = \frac{1}{nb_n} \sum_{i=1}^{n} K\left(\frac{y - Y_i}{b_n}\right), \tag{2.2}$$

for kernel function $K(\cdot)$ and bandwidth $b_n \to 0$. Under appropriate conditions, $\tilde{f}_Y(y)$ is $\sqrt{nb_n}$-consistent, and the slow convergence rate can be problematic for small sample sizes.

To overcome the aforementioned slow convergence issue, $\sqrt{n}$-consistent density estimations have been studied in the literature under two directions. One direction is $\sqrt{n}$-consistent marginal density estimation for some parametric models, including moving average process [Saavedra and Cao (1999, 2000); Schick and Wefelmeyer (2004)], invertible linear process [Schick and Wefelmeyer (2007)], nonlinear autoregressive conditional heteroscedastic model with parametric noise [Zhao (2010)], and nonlinear autoregressive model with nonparametric noise [Kim et al. (2015)]. The second direction is $\sqrt{n}$-consistent estimation for certain special integral functionals of the marginal density and/or its derivatives [Hall and Marron (1987); Bickel and Ritov (1988); Ritov and Bickel (1990); Frees (1994); Birgé and Massart (1995); Kerkyacharian and Picard (1996); Giné and Mason (2007)]. Both directions rely on either the parametric form of the model itself or explicitly known form of the functional of the marginal density, such as integrated squared density and convolution of two or more densities, with the density being the only unknown quantity. However, such parametric assumptions can be hardly justified in many applications, and there could be a risk of mis-specification.

Our first goal is to construct $\sqrt{n}$-consistent density estimation for the nonparametric regression model (2.1). Instead of imposing parametric form on $\mu(\cdot)$ as in the existing literature, we impose parametric distribution on noise e while leaving $\mu(\cdot)$ completely unspecified. Specifically, we assume

Assumption 1. In (2.1), $e \sim f_e(\cdot|\theta)$ for a density $f_e(\cdot|\theta)$ with unknown parameter θ.

Generally speaking, Assumption 1 is not as restrictive as imposing parametric model assumptions on $\mu(\cdot)$. For example, in mathematical finance, for discretely sampled data from stochastic diffusion models driven by Brownian motion, the noise is normally distributed. Furthermore, many practical data or their proper transformations satisfy (approximate) normality or, in general, the student t_d distribution with an unknown d degrees of freedom. By contrast, it can be more difficult or even unrealistic to know in advance the parametric form of $\mu(\cdot)$, especially when it is actually the goal of the study to examine the relation between variables Y and X and when such relation exhibits highly nonlinear pattern; see Models 2–4 in Section 2.3. Therefore, our nonparametric-regression-parametric-noise approach provides a more realistic and flexible framework

than those parametric-model-nonparametric-noise approaches in the literature.

In (2.1), under Assumption 1, conditioning on X, Y has density $f_e(y-\mu(X)|\theta)$. Thus,

$$f_Y(y) = \mathbb{E} f_e(y-\mu(X)|\theta). \tag{2.3}$$

Denote by $(\hat{\mu}(\cdot), \hat{\theta})$ some estimates of $(\mu(\cdot), \theta)$. By (2.3), we propose estimating $f_Y(y)$ by

$$\hat{f}_Y(y) = \frac{1}{|\mathcal{N}|} \sum_{i \in \mathcal{N}} f_e(y-\hat{\mu}(X_i)|\hat{\theta}), \quad |\mathcal{N}| \text{ is the cardinality of set } \mathcal{N}. \tag{2.4}$$

Here $\mathcal{N}$, to be specified later, is a subset of $\{1, 2, \ldots, n\}$ to avoid boundary issues in nonparametric regression. We establish the interesting result that $\hat{f}_Y(y)$ can achieve the parametric convergence rate $\sqrt{n}$, despite the nonparametric nature of the unknown function $\mu(\cdot)$ so that $\hat{\mu}(\cdot)$ is obtained through nonparametric regression method. In addition to the faster convergence rate over the nonparametric kernel density estimator $\tilde{f}_Y(y)$ in (2.2), the new estimator $\hat{f}_Y(y)$ has other appealing properties, such as being uniformly bounded, having uniformly bounded derivatives, and possessing smoothness properties which the classical estimator $\tilde{f}_Y(y)$ lacks.

The second goal is to construct $\sqrt{n}$-consistent estimator for density derivatives $f_Y^{(k)}(y) = \partial^k f_Y(y)/\partial y^k, k = 1, 2, \ldots$. The traditional nonparametric kernel density derivative estimator is to take k-th order derivative in (2.2), yielding

$$\tilde{f}_Y^{(k)}(y) = \frac{1}{n b_n^{k+1}} \sum_{i=1}^{n} K^{(k)}\left(\frac{y-Y_i}{b_n}\right), \quad \text{where} \quad K^{(k)}(u) = \frac{\partial^k K(u)}{du^k}. \tag{2.5}$$

Then $\tilde{f}_Y^{(k)}(y)$ has convergence rate $b_n^k\sqrt{nb_n}$ which deteriorates quickly as k increases, causing even more serious issue than $\tilde{f}_Y(y)$ in small to medium sample size settings. To address this, we propose estimating $f_Y^{(k)}(y)$ by taking k-th order derivative in (2.4) instead, that is,

$$\hat{f}_Y^{(k)}(y) = \frac{1}{|\mathcal{N}|} \sum_{i \in \mathcal{N}} f_e^{(k)}(y-\hat{\mu}(X_i)|\hat{\theta}) \quad \text{with} \quad f_e^{(k)}(z|\theta) = \frac{\partial f_e(z|\theta)}{\partial z^k}. \tag{2.6}$$

Similar to the case of $\hat{f}_Y(y)$, we establish $\sqrt{n}$ consistency for $\hat{f}_Y^{(k)}(y)$. As demonstrated by extensive simulation studies in Section 2.3, while $\tilde{f}_Y^{(k)}(y)$ in (2.5) can be quite unstable across realizations, $\hat{f}_Y^{(k)}(y)$ in (2.6) performs well and can reduce the mean squared error

by 95%.

In addition to providing distributional information about the data, density function has also been used in specification testing in the literature. For model (2.1), consider testing

$$H_0: \text{ in (2.1), } \mu(x) = \mu_\beta(x), \quad \text{for all } x. \tag{2.7}$$

Here, $\mu_\beta(x)$ is a given parametric model with unknown parameter β. In the literature [e.g., Aït-Sahalia (1996); Gao and King (2004); Hong and Li (2005); Aït-Sahalia et al. (2009); Zhao (2011); Kim et al. (2015)], density-based specification test measures the distance between two estimates of $f_Y(y)$: the nonparametric version $\tilde{f}_Y$ in (2.2), which is $\sqrt{nb_n}$-consistent regardless whether or not H_0 holds, and some parametric version (under H_0) which is usually $\sqrt{n}$-consistent under H_0. For example, using integrated-squared-distance (maximal deviation being another popular distance), the test is

$$D_n = \int_{\mathcal{Y}} [\tilde{f}_Y(y) - \check{f}_Y(y)]^2 dy, \quad \text{with } \tilde{f}_Y \text{ in (2.2) and a parametric estimate } \check{f}_Y. \tag{2.8}$$

Here, the construction of parametric estimate $\check{f}_Y$ is given in (2.23) below and $\mathcal{Y}$ is some bounded interval. Since $\check{f}_Y$ usually has a faster convergence rate $\sqrt{n}$ than the nonparametric estimate $\tilde{f}_Y$, the random variation in D_n is determined by the relatively poorer estimator $\tilde{f}_Y$, leading to poor performance of D_n. Pritsker (1998) pointed out that such tests [e.g., Aït-Sahalia (1996)] often require an extremely large sample size.

Our third goal is to propose more efficient density-based specification test for (2.7). In (2.8), we propose replacing the $\sqrt{nb_n}$-consistent estimator $\tilde{f}_Y$ by our newly constructed $\sqrt{n}$-consistent density estimate $\hat{f}_Y(y)$ in (2.4) to form the new test

$$T_n = \int_{\mathcal{Y}} [\hat{f}_Y(y) - \check{f}_Y(y)]^2 dy, \quad \text{with } \hat{f}_Y \text{ in (2.4) and a parametric estimate } \check{f}_Y. \tag{2.9}$$

Due to the same $\sqrt{n}$-consistency of $\hat{f}_Y(y)$ (for nonparametric $\mu(\cdot)$) and $\check{f}_Y(y)$ (under H_0), it is expected that T_n converges at rate n under H_0. Simulation studies in Section 2.3 show that $\hat{f}_Y(y)$-based test T_n performs well whereas $\tilde{f}_Y(y)$-based test D_n in (2.8) has little power.

The rest of the document is structured as follows. Section 2.2 presents large sample theory for the proposed density and density derivative estimators, simultaneous confidence band construction, and specification test. Section 2.3 contains numerical studies, and

proofs are gathered in Section 2.4.

2.2 Main results

The notation $\xrightarrow{p}$ means convergence in probability. For a matrix (or vector) $A = \{a_{ij}\}$, write $|A| = \sum_{i,j} |a_{ij}|$. For two random variables U and V, $\mathrm{Cov}(U, V)$ is their covariance.

2.2.1 $\sqrt{n}$-rate functional CLT for $\hat{f}_Y(y)$ and $\hat{f}_Y^{(k)}(y)$

For $\hat{\mu}(\cdot)$ in (2.4), we use the widely used nonparametric kernel smoothing estimate

$$\hat{\mu}(x) = \frac{\sum_{i=1}^n Y_i K\left((x - X_i)/h_n\right)}{\sum_{i=1}^n K\left((x - X_i)/h_n\right)}, \tag{2.10}$$

for bandwidth $h_n > 0$ and kernel function $K(\cdot)$.

Assumption 2. (i) $K(\cdot)$ satisfies $\int_{\mathbb{R}} K(u)du = 1$, $\int_{\mathbb{R}} uK(u)du = 0$, and has bounded support $[-L, L]$ and bounded derivative. (ii) h_n satisfies $nh_n^4 \to 0$, $\sqrt{n}h_n/\log n \to \infty$. (iii) $(X_1, Y_1, e_1), \ldots, (X_n, Y_n, e_n)$ are iid from (2.1), $\mathbb{E}(e_i) = 0$, $\mathbb{E}(e_i^2) < \infty$, and e_i is independent of X_i. (iv) X has support on a bounded interval $\mathcal{X} := [a, A]$ for some $a < A$, and the density $f_X(x)$ of X satisfies $f_X(x) > c$ for all $x \in \mathcal{X}$ for some constant $c > 0$. (v) $\mu(\cdot)$ and $f_X(\cdot)$ are twice differentiable on $\mathcal{X}$.

Remark 1. Condition (i) in Assumption 2 is typical in nonparametric kernel smoothing literature, as seen in references such as Silverman (1986) or Li and Racine (2007). We impose boundedness on the support of $K(\cdot)$ for the brevity of proofs; it may be removed at the cost of lengthier proofs. For example, Hansen (2008) established uniform convergence for nonparametric function estimation for both bounded support kernels and unbounded support kernels under some conditions on the tail of the kernel function (see Assumption 3 and the comments below it in Hansen (2008)). In particular, the Gaussian kernel, commonly used in statistical analysis [Fan and Yao (2003)], satisfies his Assumption 3. This approach is similar to that used in Fan and Yao (1998). There is a variety of kernel functions [see Silverman (1986)] that result in estimators having similar relative efficiencies, so, for example, a bounded kernel, such as the Epanechnikov kernel [see, e.g., Wand and Jones (1994)], can also be used. According to Fan and Yao (2003), it

is well-known both empirically and theoretically that the selection of kernel functions, bounded or unbounded, is not a crucial factor in the performance of the kernel estimator. As long as the kernel functions are symmetric and unimodal, the estimator's performance remains almost the same when the bandwidth h_n is optimally determined.

Under Assumption 2, it is well known [Fan and Yao (2003); Li and Racine (2007); Hansen (2008)] that

$$\sup_{x\in[a+Lh_n,A-Lh_n]} |\hat{\mu}(x) - \mu(x)| = O_p\left(h_n^2 + \sqrt{\frac{\log n}{nh_n}}\right) = o_p(n^{-1/4}). \tag{2.11}$$

Throughout, we take

$$\mathcal{N} = \{i : \ X_i \in [a + Lh_n, A - Lh_n]\}, \quad \text{see Remark 2 below.} \tag{2.12}$$

Remark 2. The introduction of $\mathcal{N}$ in (2.4) is due to pure technical consideration. As demonstrated by Li and Racine (2007), in Theorem 2.6, to ensure the uniform almost sure convergence rate of $\hat{\mu}(x)$ to $\mu(x)$ for $x \in \mathcal{X}$ in (2.11), it is necessary to exclude the boundary range of the support of X and in this way prevent boundary issues. This exclusion is achieved using the definition of $\mathcal{N}$ as given in equation (2.12). For additional details, please refer to Remark 4 in Section 2.4. In our practical implementation, we simply take $\mathcal{N} = \{1, 2, \ldots, n\}$.

As an alternative approach to avoid the boundary bias in the estimator $\hat{\mu}(x)$, it is also possible to extend the bounded support of X in Assumption 2 to unbounded support case by imposing proper decaying rate on $f_X(x) \to 0$ as $x \to \infty$. In the bounded support case, various approaches have been proposed to address the boundary or edge effects. One approach is based on boundary kernels. For example, Jones (1993) considered generalized jackknife boundary kernels for density and its derivative estimation, and Müller (1993) studied a general framework for boundary kernel construction for nonparamtric curve estimation where the kernel function includes both bounded support case and unbounded support kernels such as the Gaussian kernel. Another alternative approach [Li and Racine (2007); Fan and Yao (1998); Wand and Jones (1994)] is the local linear estimator, which possesses high statistical efficiency in an asymptotic minimax sense and it is one of the best-known approaches for boundary correction as it automatically corrects edge effects; see Fan (1992), Fan (1993), and Fan and Gijbels (1992). For a detailed treatment of the local linear estimator, refer to the monograph by Fan and Gijbels (1996). In this Chapter,

we only focus on the widely used nonparametric kernel smoothing estimate in (2.10), and it is expected that similar results can be established by using the aforementioned boundary correction methods at the cost of a lengthier proof.

For $\hat{\theta}$, since the log likelihood function for $(e_1, \ldots, e_n)$ is $\sum_{i=1}^n \log f_e(e_i|\theta)$, replacing e_i by the estimate $\hat{e}_i = Y_i - \hat{\mu}(X_i)$, we use the maximum likelihood estimate (MLE):

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \sum_{i \in \mathcal{N}} \log f_e(\hat{e}_i|\theta) \quad \text{with} \quad \hat{e}_i = Y_i - \hat{\mu}(X_i). \tag{2.13}$$

Remark 3. Instead of using the Nadaraya-Watson type kernel smoothing estimator $\hat{\mu}(x)$ in (2.10), an alternative approach is the local likelihood method. Note that, for $X_i \approx x$, $Y_i \approx \mu(x) + e_i$ has density $f_e(y - \mu(x)|\theta)$. In the first step, the local MLE of $\mu(x)$ is

$$(\tilde{\mu}(x), \tilde{\theta}) = \underset{(\mu,\theta)}{\operatorname{argmax}} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) \log f_e(Y_i - \mu|\theta). \tag{2.14}$$

In the second step, θ can be estimated by similar MLE in (2.13) with $\hat{e}_i$ therein replaced by $\tilde{e}_i = Y_i - \tilde{\mu}(X_i)$. Computationally, $\hat{\mu}(x)$ has closed-form expression while $\tilde{\mu}(x)$ has to be solved numerically subject to numerical stability issues. Furthermore, the consistency of $\hat{\mu}(x)$ only requires $\mathbb{E}(e_i) = 0$ and $\mathbb{E}(e_i^2) < \infty$, which are much less stringent than the exact distributional assumption for $\tilde{\mu}(x)$. Also, in the most important normality case, $\hat{\mu}(x)$ and $\tilde{\mu}(x)$ are equivalent. Due to these considerations, we use $\hat{\mu}(x)$.

Assumption 3. For the density $f_e(z|\theta)$ in Assumption 1, (i) $\theta \in \Theta$ for a compact set Θ. (ii) All first and second order partial derivatives of $f_e(z|\theta)$ with respect to (z, θ^T) are uniformly bounded in $z \in \mathbb{R}, \theta \in \Theta$. (iii) There exists $\epsilon > 0$ such that $\sup_{|z-e| \le \epsilon, \theta \in \Theta} |\partial \log f_e(z|\theta)/\partial z| \le \ell_1(e)$ for some function $\ell_1(\cdot)$ satisfying $\mathbb{E}\ell_1(e) < \infty$. (iv) There exists function $\ell_2(\cdot)$ such that $\sup_{\theta \in \Theta} f_e(z|\theta) \le \ell_2(z)$ and $\int_{\mathbb{R}} |z|\ell_2(z)dz < \infty$. (v) $f_e(z|\theta)$ satisfies the regularity conditions [see, e.g., Casella and Berger (2002)] for the asymptotic normality of MLE.

The regularity conditions in Assumption 3 are similar to those in the study of MLE. Some conditions may be relaxed, for example, the uniform boundedness in Assumption 3(ii) can be replaced by certain moment conditions. However, we shall not pursue this direction, since these weaker conditions come at the cost of lengthier technical proofs but without adding meaningful insights, and furthermore Assumption 3 is already flexible

enough to include most commonly used distributions, such as normal and student t distributions.

Theorem 1. *For $\hat{f}_Y(y)$ in (2.4) with $(\hat{\mu}(\cdot), \hat{\theta})$ defined in (2.10) and (2.13), under Assumptions 1–3, for any bounded interval $\mathcal{Y}$, the functional Central Limit Theorem (CLT) holds:*

$$\Big\{\sqrt{n}[\hat{f}_Y(y) - f_Y(y)], \quad y \in \mathcal{Y}\Big\} \Rightarrow \Big\{G(y), \quad y \in \mathcal{Y}\Big\}, \tag{2.15}$$

for a centered Gaussian process $\{G(y)\}$ with autocovariance $\Sigma(y, y') = \mathrm{Cov}\{G(y), G(y')\}$,

$$\begin{aligned} \Sigma(y, y') &= \Sigma_0(y, y') + \Sigma_\mu(y, y') + \Sigma_\theta(y, y'), \qquad (2.16)\\ \textit{where}\ \ \Sigma_0(y, y') &= \mathrm{Cov}\{f_e(y - \mu(X)|\theta), f_e(y' - \mu(X)|\theta)\},\\ \Sigma_\mu(y, y') &= \mathbb{E}(e^2) \times \mathbb{E}[f'_e(y - \mu(X)|\theta) f'_e(y' - \mu(X)|\theta)],\\ \Sigma_\theta(y, y') &= \mathbb{E}[\dot{f}_e(y - \mu(X)|\theta)]^T \times \mathcal{I}_\theta^{-1} \times \mathbb{E}[\dot{f}_e(y' - \mu(X)|\theta)]. \end{aligned}$$

Here, $\dot{f}_e(z|\theta) = \partial f_e(z|\theta)/\partial\theta$, $f'_e(z|\theta) = \partial f_e(z|\theta)/\partial z$, $\mathcal{I}_\theta = \mathbb{E}[\psi(e|\theta)\psi(e|\theta)^T]$ is the Fisher information matrix with the score function $\psi(e|\theta) = \partial \log f_e(e|\theta)/\partial\theta$.

By Theorem 1, for each fixed y,

$$\sqrt{n}[\hat{f}_Y(y) - f_Y(y)] \Rightarrow N\Big(0, \Sigma_0(y, y) + \Sigma_\mu(y, y) + \Sigma_\theta(y, y)\Big). \tag{2.17}$$

On the other hand, if $(\mu(\cdot), \theta)$ were known, by (2.3), we would have estimated $f_Y(y)$ by

$$f^*_Y(y) = \frac{1}{n}\sum_{i=1}^{n} f_e(y - \mu(X_i)|\theta).$$

By CLT for iid data, for each fixed y,

$$\sqrt{n}[f^*_Y(y) - f_Y(y)] \Rightarrow N\Big(0, \Sigma_0(y, y)\Big). \tag{2.18}$$

Compared to (2.18), the limiting variance in (2.17) has three components: the common term $\Sigma_0(y, y)$ accounts for the unknown distribution of X, and the two extra terms $\Sigma_\mu(y, y)$ and $\Sigma_\theta(y, y)$ reflect the variations introduced by the estimation of $\mu(\cdot)$ and θ, respectively. Furthermore, by the same argument in the proof of Theorem 1, if $\mu(\cdot)$ is known, i.e., we replace $\hat{\mu}(X_i)$ by $\mu(X_i)$ in (2.4), then the limiting variance becomes

$\Sigma_0(y,y)+\Sigma_\theta(y,y)$; if θ is known, i.e., we replace $\hat{\theta}$ by θ in (2.4), then the limiting variance becomes $\Sigma_0(y,y)+\Sigma_\mu(y,y)$.

Compared to $\sqrt{nb_n}$-consistent $\tilde{f}_Y(y)$ in (2.2), the new estimate $\hat{f}_Y(y)$ in (2.4) has a faster convergence rate $\sqrt{n}$. Intuitively, $\hat{f}_Y(y)$ uses information from the model structure which, while being nonparametric in nature, can still provide information to the density through the identity (2.3), thus leading to more efficient density estimation. Below we further present some other appealing properties of this new density estimate.

(A) Uniform boundedness.

In statistical inference, we often require the density function be bounded. From (2.4), if $f_e(z|\theta)$ is uniformly bounded, then $\hat{f}_Y(y)$ has the same uniform bound. However, for $\tilde{f}_Y(y)$ in (2.2), even though $K(\cdot)$ is bounded, there always exist sample paths (with asymptotically vanishing probability though) such that all Y_i's are close to y so that $\tilde{f}_Y(y)=O(1/b_n)$ can be arbitrarily large as $b_n\to 0$.

(B) Smoothness of the limiting process $\{G(y)\}$ in Theorem 1.

Since $\zeta_i(y)$ in (2.40) (see Section 2.4) is differentiable, $\{G(y)\}$ is differentiable. In general, if $f_e(z|\theta)$ has k-th order derivative (with respect to z), then the limiting process $\{G(y)\}$ has $(k-1)$-th order derivative. However, the nonparametric kernel density estimator $\tilde{f}_Y(y)$ in (2.2) exhibits completely different behavior. To see this, let

$$U_n(y)=\frac{\sqrt{nb_n}[\tilde{f}_Y(y)-f_Y(y)-b_n^2 f_Y''(y)\int_{\mathbb{R}}u^2K(u)du/2]}{\sqrt{f_Y(y)\int_{\mathbb{R}}K^2(u)du}}.$$

Under Assumption 2, it is well known that $U_n(y)\Rightarrow N(0,1)$ for each fixed y and that $U_n(y)$ and $U_n(y')$ are asymptotically independent $N(0,1)$ for $|y-y'|>2Lb_n$. If we take $|y-y'|=3Lb_n\to 0$, $U_n(y)-U_n(y')\Rightarrow N(0,2)$. That is, even when y and y' are arbitrarily close, the distance $U_n(y)-U_n(y')$ is a non-degenerate random variable $N(0,2)$, indicating discontinuity. Roughly speaking, this means that, for large enough n, $U_n(y)$ consists of pure jumps without any continuity, and consequently these large random variations in $\tilde{f}_Y$ causes the test D_n in (2.8) to be very insensitive to $\check{f}_Y$, or in other words, the test has no power distinguishing parametric specifications.

(C) Functional CLT.

The unpleasant non-smoothness of $\tilde{f}_Y(y)$ discussed in (B) above is caused by the lack

of tightness, which is why $\tilde{f}_Y(y)$ only has pointwise asymptotic normality and does not admit the type of functional CLT as in Theorem 1. As will be discussed in Section 2.2.2, the functional CLT is useful in studying the asymptotic distribution of the specification test.

To apply Theorem 1 for statistical inference purpose (for example, confidence interval construction), we estimate the autocovariance $\Sigma(y, y')$ in Theorem 1 by

$$\begin{aligned}
\hat{\Sigma}(y, y') &= \hat{\Sigma}_0(y, y') + \hat{\Sigma}_\mu(y, y') + \hat{\Sigma}_\theta(y, y'), \quad \text{where} \qquad (2.19)\\
\hat{\Sigma}_0(y, y') &= \frac{1}{|\mathcal{N}|}\sum_{i\in\mathcal{N}} f_e(y - \hat{\mu}(X_i)|\hat{\theta}) f_e(y' - \hat{\mu}(X_i)|\hat{\theta}) - \hat{f}_Y(y)\hat{f}_Y(y'),\\
\hat{\Sigma}_\mu(y, y') &= \frac{1}{|\mathcal{N}|}\sum_{i\in\mathcal{N}} \hat{e}_i^2 \times \frac{1}{|\mathcal{N}|}\sum_{i\in\mathcal{N}} f_e'(y - \hat{\mu}(X_i)|\hat{\theta}) f_e'(y' - \hat{\mu}(X_i)|\hat{\theta}),\\
\hat{\Sigma}_\theta(y, y') &= \frac{1}{|\mathcal{N}|}\sum_{i\in\mathcal{N}} \dot{f}_e(y - \hat{\mu}(X_i)|\hat{\theta})^T \times \left[\sum_{i\in\mathcal{N}} \psi(\hat{e}_i|\hat{\theta})\psi(\hat{e}_i|\hat{\theta})^T\right]^{-1} \times \sum_{i\in\mathcal{N}} \dot{f}_e(y' - \hat{\mu}(X_i)|\hat{\theta}).
\end{aligned}$$

Theorem 2. *Under the conditions in Theorem 1, we have uniform consistency*

$$\sup_{y,y'\in\mathcal{Y}} |\hat{\Sigma}(y, y') - \Sigma(y, y')| \xrightarrow{p} 0.$$

For the density derivative estimator $\hat{f}_Y^{(k)}(y)$ in (2.6), we have

Theorem 3. *Suppose Assumptions 1–3 hold, where Assumption 3(ii) is replaced by: All first and second order partial derivatives of $f_e(z|\theta), f_e'(z|\theta), f_e''(z|\theta), \ldots, f_e^{(k)}(z|\theta)$ with respect to (z, θ^T) are uniformly bounded in $z \in \mathbb{R}, \theta \in \Theta$. For $\hat{f}_Y^{(k)}(y)$ in (2.6), we have*

$$\left\{\sqrt{n}[\hat{f}_Y^{(k)}(y) - f_Y^{(k)}(y)], \quad y \in \mathcal{Y}\right\} \Rightarrow \left\{G^{(k)}(y), \quad y \in \mathcal{Y}\right\}, \qquad (2.20)$$

where $\{G^{(k)}(y)\}$ is the k-th order derivative of the centered Gaussian process $\{G(y)\}$ in Theorem 1. Furthermore, $\{G^{(k)}(y)\}$ is again a centered Gaussian process with autocovariance $\mathrm{Cov}\{G^{(k)}(y), G^{(k)}(y')\} = \partial^{2k}\Sigma(y, y')/\partial y^k \partial y'^k$.

By Theorem 3, all the estimated derivatives $\hat{f}_Y^{(k)}(y)$ have $\sqrt{n}$ convergence rate, whereas the nonparametric kernel density derivative estimator $\tilde{f}_Y^{(k)}(y)$ in (2.5) has a much slower convergence rate $b_n^k\sqrt{nb_n}$. Furthermore, if the derivative $f_e^{(k)}(z|\theta)$ is bounded, then $\hat{f}_Y^{(k)}(y)$ has the same bound. However, for $\tilde{f}_Y^{(k)}(y)$ in (2.5), as argued in (A) above, there exist

sample paths such that $\tilde{f}_Y^{(k)}(y)$ grows at rate $O(1/b_n^{k+1})$, which causes the estimated curve quite unstable. The simulation studies in Section 2.3 confirm this phenomenon, and $\hat{f}'_Y(y)$ can reduce the mean squared error by 95% compared to $\tilde{f}'_Y(y)$.

Bandwidth selection: For the bandwidth h_n in (2.10), if we let $h_n \propto n^{-\kappa}$, then Assumption 2(ii) holds for $\kappa \in (1/4, 1/2)$. On the other hand, for nonparametric mean regression function estimation, it is well known that Ruppert et al. (1995)'s automatic plug-in bandwidth selector (implemented via the `dpill` function in `R`) selects the optimal bandwidth $h_n^* \propto n^{-1/5}$. Thus, a reasonable choice of h_n is

$$h_n = n^{-\gamma} h_n^* \text{ for some } \gamma, \quad \text{where } h_n^* \text{ is automatic plug-in bandwidth (dpill in R).} \tag{2.21}$$

From $h_n^* \propto n^{-1/5}$, $h_n \propto n^{-(1/5+\gamma)}$. From $1/5 + \gamma \in (1/4, 1/2)$, we have $\gamma \in (1/20, 3/10)$. By the simulation studies in Section 2.3, the proposed method is quite robust against the choice of γ, and in practice we can take $\gamma = 7/40$, the middle point of $(1/20, 3/10)$.

2.2.2 Specification test

The $\sqrt{n}$-consistent density estimate $\hat{f}_Y(y)$ can be used to address specification testing problem (2.7) through the test T_n in (2.9). We need to construct the parametric estimate $\check{f}_Y$ of $f_Y(y)$. Under H_0, we estimate β by the nonlinear least-squares method:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} [Y_i - \mu_\beta(X_i)]^2. \tag{2.22}$$

For θ, we use the same estimate $\hat{\theta}$ in (2.13) where the nonparametric estimate $\hat{\mu}(\cdot)$ is used to obtain $\hat{e}_i$, due to the consideration that $\hat{\theta}$ is always $\sqrt{n}$-consistent, regardless of whether or not H_0 holds. Therefore, similar to (2.4), under H_0, the parametric estimate of $f_Y(y)$ is

$$\check{f}_Y(y) = \frac{1}{n} \sum_{i=1}^{n} f_e(y - \mu_{\hat{\beta}}(X_i) | \hat{\theta}), \quad \text{with } \hat{\theta} \text{ in (2.13).} \tag{2.23}$$

The proposed test T_n is then obtained by plugging $\hat{f}_Y$ and $\check{f}_Y$ into (2.9).

Assumption 4. (i) $\beta \in \Omega$ for a compact set Ω. (ii) $\mathbb{P}\{\mu_{\beta_1}(X) = \mu_{\beta_2}(X)\} = 1$ implies $\beta_1 = \beta_2$. (iii) For each $\beta \in \Omega$, $\mu_\beta(x)$ is twice differentiable in $x \in \mathcal{X}$ ($\mathcal{X}$ is defined in

Assumption 2). (iv) For each $x \in \mathcal{X}$, the partial derivatives $\dot{\mu}_\beta(x) = \partial\mu_\beta(x)/\partial\beta$ and $\ddot{\mu}_\beta(x) = \partial^2\mu_\beta(x)/\partial\beta\partial\beta^T$ are continuous in $\beta \in \Omega$. (v) $\mathbb{E}[\dot{\mu}_\beta(X)\dot{\mu}_\beta(X)^T]$ is non-singular.

Theorem 4. *Suppose Assumptions 1–4 hold. Write* $\dot{\mu}_\beta(X) = \partial\mu_\beta(X)/\partial\beta, M(y) = \mathbb{E}[f_e'(y-\mu_\beta(X)|\theta)\dot{\mu}_\beta(X)^T]$, $\mathcal{J}_\beta = \mathbb{E}[\dot{\mu}_\beta(X)\dot{\mu}_\beta(X)^T]$. *For* T_n *in (2.9) with* $\hat{f}_Y$ *in (2.4) and* $\check{f}_Y$ *in (2.23),*

$$nT_n \Rightarrow \int_{\mathcal{Y}} Z(y)^2 dy, \quad \textit{under } H_0,$$

where $\{Z(y)\}$ *is a centered Gaussian process with autocovariance*

$$\mathrm{Cov}\{Z(y), Z(y')\} = \mathbb{E}(e^2)\Big\{\mathbb{E}[f_e'(y-\mu_\beta(X)|\theta)f_e'(y'-\mu_\beta(X)|\theta)] - M(y)\mathcal{J}_\beta^{-1}M(y')^T\Big\}.$$

Since the asymptotic distribution in Theorem 4 involves complicated unknown functions, we adopt bootstrap approach to obtain the cutoff value at significance level α:

(1) Based on the original data $(X_1, Y_1), \ldots, (X_n, Y_n)$ and given interval $\mathcal{Y}$, obtain h_n in (2.21) with $\gamma = 7/40$ [see the discussion below (2.21)], calculate $\hat{\mu}(\cdot)$ in (2.10), $\hat{\theta}$ in (2.13), $\hat{\beta}$ in (2.22), and further calculate T_n in (2.9) with $\hat{f}_Y$ in (2.4) and $\check{f}_Y$ in (2.23).

(2) Using $\hat{\mu}(\cdot)$ and $\hat{\beta}$ in step (1) to obtain $\hat{e}_i = Y_i - \hat{\mu}(X_i)$ and $\mu_{\hat{\beta}}(X_i), i = 1, \ldots, n$.

(3) Draw n samples, denoted by $\{\hat{e}_1^b, \ldots, \hat{e}_n^b\}$, from $\{\hat{e}_1, \ldots, \hat{e}_n\}$ with replacement, and calculate the test statistic based on the bootstrap data $(X_1, Y_1^b), \ldots, (X_n, Y_n^b)$, where $Y_i^b = \mu_{\hat{\beta}}(X_i) + \hat{e}_i^b$, using the same interval $\mathcal{Y}$ and bandwidth h_n in step (1).

(4) Repeat step (3) a large number (say, M) of times to obtain M realizations of the test statistic, and reject H_0 if T_n exceeds the $(1-\alpha)$ quantile of these realizations.

2.2.3 The special normality case

The normality assumption is one of the most common assumptions in statistical analysis and can be satisfied for many practical data after proper transformations. Below we consider the special normality case of (2.1):

$$Y = \mu(X) + e, \quad e \sim N(0, \theta) \text{ with unknown } \theta > 0. \tag{2.24}$$

Denote by $\phi(y;\mu,\theta)$ the normal density of $N(\mu,\theta)$, and denote the partial derivatives by $\dot{\phi}_\mu(y;\mu,\theta) = \partial\phi(y;\mu,\theta)/\partial\mu$ and $\dot{\phi}_\theta(y;\mu,\theta) = \partial\phi(y;\mu,\theta)/\partial\theta$. Then (2.4) becomes

$$\hat{f}_Y(y) = \frac{1}{|\mathcal{N}|}\sum_{i\in\mathcal{N}} \phi(y;\hat{\mu}(X_i),\hat{\theta}) \quad \text{with} \quad \hat{\theta} = \frac{1}{|\mathcal{N}|}\sum_{i\in\mathcal{N}}[Y_i - \hat{\mu}(X_i)]^2, \quad \hat{\mu}(\cdot) \text{ in (2.10).} \tag{2.25}$$

From Theorem 1, after some calculations we can obtain

Corollary 1. *For model (2.24), under Assumption 2, the functional CLT in Theorem 1 holds with the autocavariance $\Sigma(y,y')$ therein reducing to*

$$\begin{aligned}\Sigma(y,y') &= \text{Cov}\{\phi(y;\mu(X),\theta),\phi(y';\mu(X),\theta)\} + \theta \times \mathbb{E}[\dot{\phi}_\mu(y;\mu(X),\theta)\dot{\phi}_\mu(y';\mu(X),\theta)] \\ &\quad +2\theta^2 \times \mathbb{E}[\dot{\phi}_\theta(y;\mu(X),\theta)] \times \mathbb{E}[\dot{\phi}_\theta(y';\mu(X),\theta)].\end{aligned}$$

As in Theorem 2, a uniformly consistent estimate of $\Sigma(y,y')$ is given by

$$\begin{aligned}\hat{\Sigma}(y,y') &= \frac{1}{|\mathcal{N}|}\sum_{i\in\mathcal{N}} \phi(y;\hat{\mu}(X_i),\hat{\theta})\phi(y';\hat{\mu}(X_i),\hat{\theta}) - \hat{f}_Y(y)\hat{f}_Y(y') \\ &\quad +\hat{\theta} \times \frac{1}{|\mathcal{N}|}\sum_{i\in\mathcal{N}} \dot{\phi}_\mu(y;\hat{\mu}(X_i),\hat{\theta})\phi_\mu(y';\hat{\mu}(X_i),\hat{\theta}) \\ &\quad +2\hat{\theta}^2 \times \frac{1}{|\mathcal{N}|}\sum_{i\in\mathcal{N}} \dot{\phi}_\theta(y;\hat{\mu}(X_i),\hat{\theta}) \times \frac{1}{|\mathcal{N}|}\sum_{i\in\mathcal{N}} \dot{\phi}_\theta(y';\hat{\mu}(X_i),\hat{\theta}).\end{aligned} \tag{2.26}$$

From Theorem 1 and the discussions in Section 2.2.1, in the presence of unknown nonparametric mean regression function $\mu(\cdot)$ in (2.1), the proposed density estimator can achieve $\sqrt{n}$ parametric convergence rate and possesses several appealing properties. A related question is how the method performs in the absence of such nonparametric mean function. Below we answer this question for the normality case (2.24).

Consider the special case of (2.1) where $\mu(\cdot) \equiv \mu$ is actually an unknown constant μ:

$$Y = \mu + e, \quad \mu \text{ is unknown constant, } e \sim N(0,\theta) \text{ with unknown } \theta > 0. \tag{2.27}$$

However, without knowing that $\mu(\cdot) \equiv \mu$ is indeed a constant, we still use $\hat{f}_Y(y)$ in (2.25) to estimate $f_Y(y)$. Theorem 5 below establishes that $\hat{f}_Y(y)$ is asymptotically efficient in the sense that it is asymptotically equivalent to the MLE obtained as if we have known that $\mu(\cdot) \equiv \mu$ is a constant. That is, in the absence of nonparametric mean function, the

proposed method will not lose any efficiency by fitting nonparametric regression.

Theorem 5. *For the constant mean model (2.27), denote by* $\hat{f}_Y^{\text{MLE}}(y) = \phi(y; \overline{\mu}, \overline{\theta})$ *the MLE of* $f_Y(y)$*, where* $\overline{\mu} = n^{-1}\sum_{i=1}^n Y_i$ *and* $\overline{\theta} = n^{-1}\sum_{i=1}^n (Y_i - \overline{\mu})^2$ *are the MLE of* (μ, θ)*. For* $\hat{f}_Y(y)$ *in (2.25), under the conditions in Theorem 1, for any bounded interval* $\mathcal{Y}$*, we have*

$$\sup_{y\in\mathcal{Y}} |\hat{f}_Y(y) - \hat{f}_Y^{\text{MLE}}(y)| = o_p(n^{-1/2}).$$

2.3 Numerical studies

Our numerical studies have four goals:

(i) Comparing the new density estimate $\hat{f}_Y(y)$ in (2.4) to the nonparametric kernel density estimator $\tilde{f}_Y(y)$ in (2.2) which is implemented using `R` function `density` with the default rule-of-thumb bandwidth [Silverman (1986)] choice for b_n (see Subsection 2.3.1).

(ii) Comparing the new density derivative estimate $\hat{f}_Y^{(k)}(y)$ in (2.6) to the nonparametric kernel density derivative estimator $\tilde{f}_Y^{(k)}(y)$ in (2.5) which is implemented by `dkde` function in `R` package `kedd` with the default unbiased cross-validation choice for b_n (see Subsection 2.3.2).

(iii) Examining the effect of sample size n, bandwidth h_n, and different models on the performance of the proposed density and density derivative estimates.

(iv) Examining the performance of the specification test T_n in (2.9) (see Subsection 2.3.3).

Consider $(X_1, Y_1), \ldots, (X_n, Y_n)$ from four increasingly more complicated models

$$\begin{aligned}
&\text{Model 1:} \quad Y = 0.3 + 0.4X + e;\\
&\text{Model 2:} \quad Y = 0.3 + 0.4\sin(2\pi X) + e;\\
&\text{Model 3:} \quad Y = 0.3 + 0.4X + 0.4\sin(2\pi X) + e;\\
&\text{Model 4:} \quad Y = 0.3 + 0.4X + 0.4\sin(2\pi X) + 0.5\sqrt{1+X^2} + e.
\end{aligned}$$

For X, we use $X \sim \text{Unif}(0, 1)$. For noise e, we consider two most widely used distributions:

(i) $e \sim N(0, 0.6^2)$;

(ii) $e \sim 0.6t_4$, student t with 4 degrees of freedom and scale parameter 0.6.

The second distribution is a heavy-tailed distribution with infinite fourth moment and infinite kurtosis. The true density $f_Y(y)$ of Y is computed by numerically evaluating (2.3).

2.3.1 Comparison of density estimations

To examine the effect of bandwidth h_n in (2.10), we consider four choices of $\gamma \in (1/20, 3/10) : 2/20, 3/20, 4/20, 5/20$ in (2.21); see the discussions below (2.21). Also, four sample sizes are used $n = 50, 100, 200, 400$, ranging from small to medium sample sizes. For $\tilde{f}_Y(y)$, we compute its mean-squared-error (MSE) as

$$\text{MSE}(\tilde{f}_Y) = \text{average of 1000 realizations of } \left\{ \frac{q_Y(0.975) - q_Y(0.025)}{100} \sum_{j=1}^{100} [\tilde{f}_Y(y_j) - f_Y(y_j)]^2 \right\},$$

where $q_Y(0.025)$ and $q_Y(0.975)$ are the 2.5% and 97.5% quantile of Y, respectively, and $y_1, \ldots, y_{100}$ are equally spaced points on $[q_Y(0.025), q_Y(0.975)]$. The MSE for $\hat{f}_Y(y)$ is defined similarly. For ease of comparison, we also compute the relative MSE (RMSE) of $\hat{f}_Y$ as $\text{MSE}(\hat{f}_Y)/\text{MSE}(\tilde{f}_Y)$, with RMSE ≤ 1 indicating better performance of $\hat{f}_Y(y)$.

MSE and RMSE of $\hat{f}_Y$ are presented in Table 2.1. We make the following observations:

(1^0) As sample size increases, MSE decreases quickly, indicating increasingly better performance.

(2^0) The proposed density estimate $\hat{f}_Y(y)$ has substantial reductions in MSE with RMSE below 70% for most cases. For example, for $n = 100$, the reduction in MSE is about 50% (respectively, 35%) for normal (respectively, student t) distribution.

(3^0) The larger the sample size, the smaller the RMSE, and the greater the relative advantage of $\hat{f}_Y(y)$ over $\tilde{f}_Y(y)$. Both $\hat{f}_Y(y)$ and $\tilde{f}_Y(y)$ have bias terms which, while being theoretically negligible, may have some effect in finite samples, and these bias terms decrease as sample size increases. Since MSE = Bias^2 + Variance, as bias diminishes, the variance plays a dominating role, leading to increasingly better

relative performance of $\sqrt{n}$-consistent $\hat{f}_Y(y)$ than $\sqrt{nb_n}$-consistent $\tilde{f}_Y(y)$, where $b_n \propto n^{-1/5}$ decreases with sample size.

(4^0) The performance of $\hat{f}_Y(y)$ is consistent across the choices of bandwidth h_n through the choices of γ in (2.21). Thus, in practice, a reasonable and safe choice of γ is 7/40, the middle point of $(1/20, 3/10)$.

(5^0) The relative advantage of $\hat{f}_Y(y)$ over $\tilde{f}_Y(y)$ is consistent across the four models which range from simple linear model to highly nonlinear model. In general, the relative advantage is smaller for models with student t_4 noise, which is not surprising as this distribution has a very heavy tail with infinite fourth moment.

Table 2.1. MSE of $\hat{f}_Y$ and relative MSE defined as RMSE $= \frac{\text{MSE}(\hat{f}_Y)}{\text{MSE}(\tilde{f}_Y)}$, using bandwidth h_n in (2.21) with $\gamma = 2/20, 3/20, 4/20, 5/20$. RMSE ≤ 1 indicates better performance of $\hat{f}_Y(y)$.

			$\gamma = 2/20$		$\gamma = 3/20$		$\gamma = 4/20$		$\gamma = 5/20$	
		n	MSE	RMSE	MSE	RMSE	MSE	RMSE	MSE	RMSE
Model 1	$e \sim N(0, 0.6^2)$	50	0.0106	0.614	0.0104	0.609	0.0106	0.636	0.0113	0.653
		100	0.0047	0.479	0.0045	0.468	0.0047	0.492	0.0050	0.514
		200	0.0021	0.380	0.0021	0.373	0.0023	0.408	0.0025	0.419
		400	0.0010	0.277	0.0010	0.282	0.0010	0.306	0.0011	0.316
	$e \sim 0.6t_4$	50	0.0122	0.751	0.0124	0.738	0.0121	0.744	0.0130	0.754
		100	0.0060	0.641	0.0063	0.673	0.0071	0.714	0.0061	0.635
		200	0.0033	0.573	0.0032	0.558	0.0032	0.564	0.0034	0.602
		400	0.0016	0.465	0.0017	0.505	0.0017	0.502	0.0017	0.544
Model 2	$e \sim N(0, 0.6^2)$	50	0.0098	0.642	0.0103	0.670	0.0102	0.648	0.0114	0.738
		100	0.0044	0.491	0.0045	0.503	0.0046	0.539	0.0049	0.546
		200	0.0020	0.377	0.0021	0.397	0.0022	0.410	0.0023	0.428
		400	0.0010	0.326	0.0010	0.307	0.0011	0.342	0.0010	0.338
	$e \sim 0.6t_4$	50	0.0114	0.776	0.0115	0.794	0.0115	0.758	0.0126	0.811
		100	0.0059	0.683	0.0056	0.655	0.0059	0.694	0.0055	0.656
		200	0.0027	0.558	0.0028	0.559	0.0031	0.612	0.0029	0.575
		400	0.0015	0.483	0.0014	0.469	0.0015	0.507	0.0015	0.525
Model 3	$e \sim N(0, 0.6^2)$	50	0.0098	0.585	0.0106	0.652	0.0105	0.626	0.0115	0.698
		100	0.0045	0.480	0.0045	0.479	0.0046	0.478	0.0054	0.554
		200	0.0021	0.387	0.0022	0.393	0.0022	0.405	0.0042	0.419
		400	0.0009	0.296	0.0010	0.305	0.0010	0.304	0.0011	0.319
	$e \sim 0.6t_4$	50	0.0114	0.704	0.0116	0.733	0.0116	0.774	0.0121	0.776
		100	0.0058	0.631	0.0062	0.703	0.0057	0.636	0.0061	0.663
		200	0.0031	0.591	0.0031	0.569	0.0032	0.593	0.0032	0.592
		400	0.0016	0.479	0.0016	0.489	0.0015	0.470	0.0016	0.509
Model 4	$e \sim N(0, 0.6^2)$	50	0.0095	0.575	0.0114	0.659	0.0106	0.653	0.0114	0.684
		100	0.0046	0.478	0.0045	0.468	0.0047	0.496	0.0050	0.534
		200	0.0021	0.378	0.0023	0.395	0.0023	0.413	0.0024	0.429
		400	0.0010	0.316	0.0010	0.317	0.0011	0.327	0.0011	0.340
	$e \sim 0.6t_3$	50	0.0120	0.753	0.0115	0.738	0.0116	0.733	0.0126	0.776
		100	0.0062	0.655	0.0060	0.641	0.0065	0.679	0.0061	0.675
		200	0.0031	0.576	0.0030	0.571	0.0031	0.582	0.0033	0.597
		400	0.0016	0.504	0.0016	0.492	0.0016	0.498	0.0015	0.470

2.3.2 Comparison of density derivative estimations

Table 2.2 presents the MSE of the new density derivative estimate $\hat{f}'_Y(y)$ in (2.6) and its RMSE to the nonparametric kernel density derivative estimator $\tilde{f}'_Y(y)$ in (2.5). Compared to the density estimation case, the reduction in MSE of density derivative estimation is even more striking and exceeds 90% (i.e., RMSE $\leq 10\%$) in all cases considered. Since $\hat{f}'_Y(y)$ is $\sqrt{n}$-consistent while $\tilde{f}'_Y(y)$ is only $b_n\sqrt{nb_n}$-consistent, the larger convergence rate ratio $b_n^{-1.5}$ explains the significantly larger reductions in MSE compared to the density estimation case where the convergence rate ratio is $b_n^{-0.5}$.

Table 2.2. MSE of the derivative estimate $\hat{f}'_Y$ and relative MSE defined as RMSE $= \frac{\text{MSE}(\hat{f}'_Y)}{\text{MSE}(\tilde{f}'_Y)}$, using bandwidth h_n in (2.21) with $\gamma = 2/20, 3/20, 4/20, 5/20$. RMSE ≤ 1 indicates better performance of $\hat{f}'_Y(y)$.

			$\gamma = 2/20$		$\gamma = 3/20$		$\gamma = 4/20$		$\gamma = 5/20$	
		n	MSE	RMSE	MSE	RMSE	MSE	RMSE	MSE	RMSE
Model 1	$e \sim N(0, 0.6^2)$	50	0.0633	0.038	0.0626	0.046	0.0689	0.047	0.0784	0.053
		100	0.0249	0.037	0.0256	0.029	0.0299	0.046	0.0312	0.037
		200	0.0111	0.016	0.0109	0.020	0.0123	0.023	0.0127	0.022
		400	0.0051	0.017	0.0051	0.013	0.0055	0.014	0.0060	0.020
	$e \sim 0.6t_4$	50	0.0694	0.078	0.0642	0.093	0.0672	0.099	0.0695	0.084
		100	0.0356	0.073	0.0365	0.058	0.0360	0.054	0.0371	0.056
		200	0.0187	0.047	0.0187	0.068	0.0196	0.059	0.0201	0.053
		400	0.0100	0.048	0.0101	0.046	0.0103	0.044	0.0101	0.049
Model 2	$e \sim N(0, 0.6^2)$	50	0.0493	0.040	0.0569	0.062	0.0593	0.062	0.0666	0.062
		100	0.0212	0.030	0.0236	0.035	0.0242	0.030	0.0251	0.033
		200	0.0098	0.020	0.0099	0.025	0.0106	0.029	0.0113	0.022
		400	0.0044	0.015	0.0048	0.015	0.0049	0.019	0.0050	0.019
	$e \sim 0.6t_4$	50	0.0546	0.067	0.0568	0.061	0.0573	0.094	0.0635	0.098
		100	0.0269	0.074	0.0291	0.064	0.0282	0.066	0.0307	0.065
		200	0.0142	0.043	0.0153	0.044	0.0141	0.051	0.0150	0.060
		400	0.0069	0.034	0.0076	0.041	0.0727	0.042	0.0747	0.039
Model 3	$e \sim N(0, 0.6^2)$	50	0.0528	0.033	0.0621	0.049	0.0656	0.052	0.0678	0.047
		100	0.0232	0.030	0.0239	0.042	0.0268	0.037	0.0278	0.038
		200	0.0108	0.023	0.0118	0.025	0.0170	0.026	0.0123	0.025
		400	0.0051	0.018	0.0051	0.016	0.0053	0.013	0.0056	0.019
	$e \sim 0.6t_4$	50	0.0632	0.0692	0.0612	0.085	0.0637	0.082	0.0665	0.072
		100	0.0310	0.059	0.0371	0.071	0.0322	0.057	0.0320	0.061
		200	0.0160	0.038	0.0171	0.052	0.0170	0.066	0.0172	0.050
		400	0.0089	0.037	0.0087	0.039	0.0088	0.051	0.0907	0.043
Model 4	$e \sim N(0, 0.6^2)$	50	0.0553	0.030	0.0623	0.063	0.0669	0.060	0.0773	0.057
		100	0.0259	0.032	0.0269	0.027	0.0252	0.036	0.0294	0.030
		200	0.0109	0.018	0.0114	0.023	0.0125	0.023	0.0132	0.028
		400	0.0053	0.013	0.0055	0.014	0.0054	0.017	0.0060	0.014
	$e \sim 0.6t_3$	50	0.0668	0.082	0.0709	0.083	0.0697	0.089	0.0677	0.089
		100	0.0324	0.061	0.0331	0.057	0.0346	0.068	0.0352	0.068
		200	0.0174	0.044	0.0161	0.059	0.0184	0.051	0.0184	0.046
		400	0.0886	0.041	0.0856	0.030	0.0927	0.043	0.0096	0.044

From the discussion below Theorem 3, the nonparametric kernel density derivative estimator $\tilde{f}'_Y(y)$ in (2.5) can be quite unstable. To illustrate this, Figure 2.1 plots two realizations of density estimations (left panel) and density derivative estimations (right panel) from Model 2 with $e \sim 0.6t_4$ and $n = 400$. Using $\gamma = 7/40$ in (2.21), $\hat{f}_Y(y)$ (dashed curve in left panel) and $\hat{f}'_Y(y)$ (dashed curve in right panel) match the true curves (solid curve in all plots) very well for both realizations. By contrast, while $\tilde{f}_Y(y)$ (dotted curve in left panel) works reasonably well for density estimate (not as good as $\hat{f}_Y(y)$ though), $\tilde{f}'_Y(y)$ (dotted curve in right panel) has very poor performance in the second realization. In our simulations, realizations with such poor performance of $\tilde{f}'_Y(y)$ occur quite often, resulting in MSE($\tilde{f}'_Y$) being about 20 times larger than MSE($\hat{f}'_Y$).

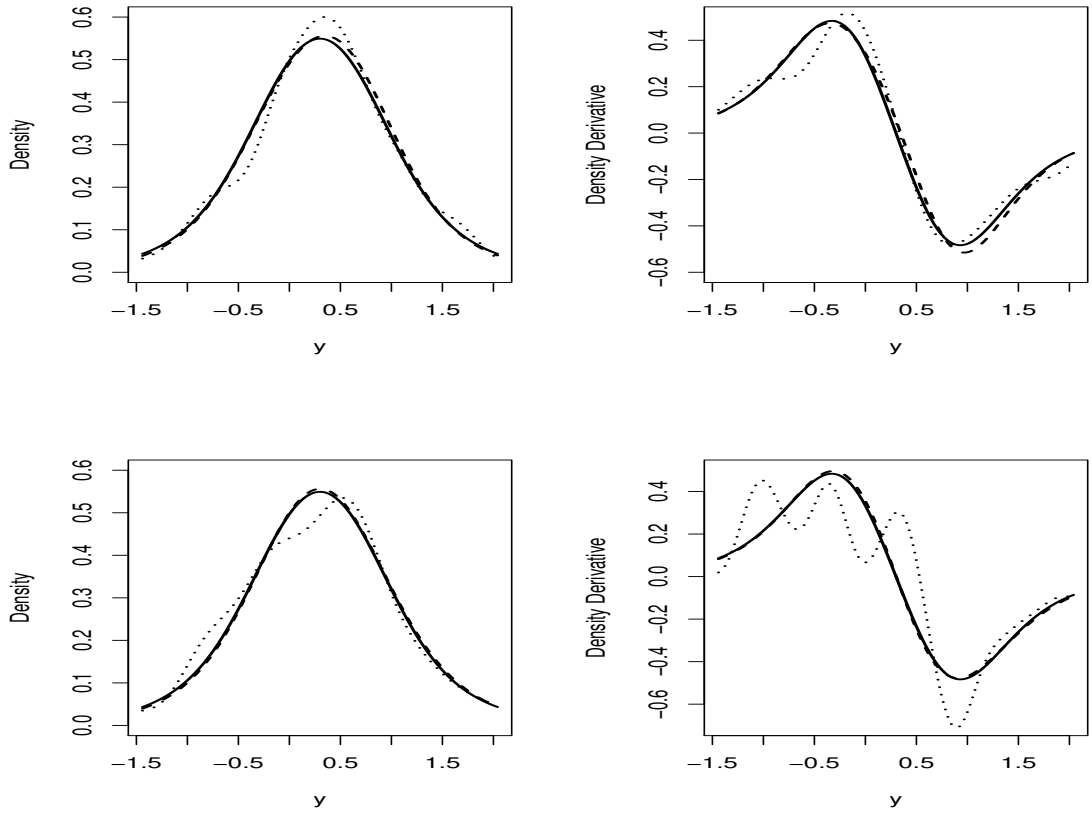

Figure 2.1. Two realizations (one realization on each row) of density estimates (left plots) and density derivative estimates (right plots) from Model 2 with $e \sim 0.6t_4$ and $n = 400$. In each plot, solid curve (——) is the true function, dotted curve ($\cdots$) is the nonparametric kernel estimates, and dashed curve ($--$) is the proposed method.

2.3.3 Comparison of specification tests T_n and D_n

In this Subsection, we compare specification tests T_n in (2.9) and D_n in (2.8). Consider testing $H_0 : \mu(x) = \beta_0 + \beta_1 x$ when the data are actually generated from Model 1 (null model) and Models 2–4 (alternative models) with noise $e \sim N(0, 0.6^2)$. As discussed in Section 2.2.2, the cutoff value of T_n is obtained by bootstrap method with 1000 bootstrap realizations and, for fair comparison, the cutoff value of D_n is obtained by similar bootstrap method. Using $\gamma = 7/40$ in (2.21), Table 2.3 summarizes the empirical size (Model 1) and empirical power (Models 2–4) based on 100 realizations, for sample size $n = 50, 100, 200, 400$, and at significance level $\alpha = 5\%$. Clearly, test T_n has empirical size close to the nominal level and its power increases drastically with sample size. At $n = 200$, T_n has power about 85%; at $n = 400$, T_n has power 100%. By contrast, test D_n has little power detecting deviations from null hypothesis, and this poor performance is consistent with the findings in Pritsker (1998) and Kim et al. (2015). In our simulations we also tried significance levels 10% and 1% and the conclusions are similar.

Table 2.3. Comparison of size (under null Model 1) and power (under alternative Models 2–4) for specification tests T_n (the first number in each cell) and D_n (the bracketed number), significance level 5%.

		$n = 50$	$n = 100$	$n = 200$	$n = 400$
Null model:	Model 1	0.04(0.01)	0.03(0.01)	0.05(0.00)	0.04(0.00)
Alternative models:	Model 2	0.18(0.03)	0.45(0.06)	0.88(0.09)	1.00(0.09)
	Model 3	0.25(0.04)	0.51(0.03)	0.85(0.04)	1.00(0.03)
	Model 4	0.19(0.02)	0.49(0.03)	0.84(0.02)	1.00(0.04)

2.4 Proofs

Write $\|Z\|^2 = \mathbb{E}(Z^2)$ for a random variable Z. Throughout, $c_1, c_2, \ldots$, are generic constants that may vary from places to places. Also, $\mathcal{Y}$ is any bounded interval with length denoted by $|\mathcal{Y}|$. First, we present two useful lemmas.

Lemma 1. *(i) For any differentiable random function $H(y)$,*

$$\sup_{y\in\mathcal{Y}} |H(y)| = O_p\left\{\inf_{y\in\mathcal{Y}} \|H(y)\| + \sup_{y\in\mathcal{Y}} \|H'(y)\|\right\}. \tag{2.28}$$

(ii) If $H(y) = \sum_{i=1}^n H_i(y)$*, where* $H_1(y), \ldots, H_n(y)$ *are iid and differentiable, then*

$$\sup_{y\in\mathcal{Y}} |H(y) - n\mathbb{E}H_1(y)| = O_p\left\{\sqrt{n}\left[\inf_{y\in\mathcal{Y}} \|H_1(y)\| + \sup_{y\in\mathcal{Y}} \|H_1'(y)\|\right]\right\}. \tag{2.29}$$

Proof. (i) For any $y, y_0 \in \mathcal{Y}$, $|H(y) - H(y_0)| = |\int_{y_0}^y H'(z)dz| \le \int_{\mathcal{Y}} |H'(z)|dz$. Thus,

$$\sup_{y\in\mathcal{Y}}[H(y) - H(y_0)]^2 \le \left[\int_{\mathcal{Y}} |H'(z)|dz\right]^2 \le |\mathcal{Y}| \int_{\mathcal{Y}} H'(z)^2 dz,$$

where the second "$\le$" follows from Cauchy-Schwarz inequality. Taking expectation yields

$$\mathbb{E}\left\{\sup_{y\in\mathcal{Y}}[H(y) - H(y_0)]^2\right\} \le |\mathcal{Y}| \int_{\mathcal{Y}} \mathbb{E}[H'(z)^2]dz \le |\mathcal{Y}|^2 \sup_{y\in\mathcal{Y}} \mathbb{E}[H'(y)^2] = O(1) \sup_{y\in\mathcal{Y}} \mathbb{E}[H'(y)^2],$$

which, after taking square root and using triangle inequality, implies

$$\left\|\sup_{y\in\mathcal{Y}} |H(y)|\right\| = O(1)\left[\|H(y_0)\| + \sup_{y\in\mathcal{Y}} \|H'(y)\|\right].$$

Since y_0 is arbitrary, (2.28) follows. (ii) By writing $H(y) - n\mathbb{E}H_1(y) = \sum_{i=1}^n [H_i(y) - \mathbb{E}H_i(y)]$ with the summands being iid with zero mean, the desired result follows from (2.28). ♢

The following lemma can greatly simplify the presentation of the main proofs.

Lemma 2. *Let* $g(\cdot)$ *and* $h(\cdot,\cdot)$ *be functions such that* $\mathbb{E}[g(e_i)] = 0$, $\mathbb{E}[g^2(e_i)] < \infty$, $\sup_{y\in\mathcal{Y}} \mathbb{E}\{h^2(y, X_i) + [\partial h(y, X_i)/\partial y]^2\} < \infty$. *For* $\mathcal{N}$ *in (2.12), define*

$$\Lambda_n(y) = \frac{1}{|\mathcal{N}|} \sum_{i\in\mathcal{N}} g(e_i)h(y, X_i) - \frac{1}{n}\sum_{i=1}^n g(e_i)h(y, X_i).$$

Then $\sup_{y\in\mathcal{Y}} |\Lambda_n(y)| = o_p(n^{-1/2})$.

Proof. Let $\mathcal{N}^c$ be the complement of $\mathcal{N}$ and $\mathbf{1}$ the indicator function. Introducing

$$\Lambda_{n,1}(y) = \sum_{i=1}^n g(e_i)h(y, X_i) \quad \text{and} \quad \Lambda_{n,2}(y) = \sum_{i=1}^n g(e_i)h(y, X_i)\mathbf{1}_{i\in\mathcal{N}^c},$$

we can rewrite

$$\Lambda_n(y) = \left(\frac{1}{|\mathcal{N}|} - \frac{1}{n}\right)\Lambda_{n,1}(y) - \frac{1}{|\mathcal{N}|}\Lambda_{n,2}(y). \tag{2.30}$$

By Assumption 2(iv), $|\mathcal{N}^c| = O_p(nh_n) = o_p(n)$, which implies $1/|\mathcal{N}| - 1/n = o_p(1/n)$. By (2.30), to prove $\sup_{y\in\mathcal{Y}} |\Lambda_n(y)| = o_p(n^{-1/2})$, it suffices to prove

$$\sup_{y\in\mathcal{Y}} |\Lambda_{n,1}(y)| = O_p(\sqrt{n}) \quad \text{and} \quad \sup_{y\in\mathcal{Y}} |\Lambda_{n,2}(y)| = o_p(\sqrt{n}). \tag{2.31}$$

Since $\{(e_i, X_i)\}$ are iid and e_i is independent of X_i, the assumption $\mathbb{E}[g(e_i)] = 0$ implies that the summands in $\Lambda_{n,1}(y)$ are iid and have zero mean. By the same argument, the summands in the derivative $\Lambda'_{n,1}(y)$ are also iid and have zero mean. Therefore, an application of Lemma 1(ii) gives $\sup_{y\in\mathcal{Y}} |\Lambda_{n,1}(y)| = O_p(\sqrt{n})$. For $\Lambda_{n,2}(y)$, under the given conditions, by the Dominated Convergence Theorem,

$$\|g(e_i)h(y, X_i)\mathbf{1}_{i\in\mathcal{N}^c}\|^2 \to 0 \quad \text{and} \quad \left\|g(e_i)\frac{\partial h(y, X_i)}{\partial y}\mathbf{1}_{i\in\mathcal{N}^c}\right\|^2 \to 0.$$

Then an application of Lemma 1(ii) gives $\sup_{y\in\mathcal{Y}} |\Lambda_{n,2}(y)| = o_p(\sqrt{n})$. ♢

Remark 4. Thanks to the introduction of $\mathcal{N}$ in (2.12), by (2.11), we have

$$\max_{i\in\mathcal{N}} |\hat{\mu}(X_i) - \mu(X_i)| = o_p(n^{-1/4}). \tag{2.32}$$

A careful examination of the proofs below reveals that the only reason we need to impose $i \in \mathcal{N}$ is to ensure the uniform approximation (2.32). In subsequent proofs, for ease of presentation, we shall pretend that the uniform approximation (2.32) holds for all i, that is,

$$\max_{1\le i\le n} |\hat{\mu}(X_i) - \mu(X_i)| = o_p(n^{-1/4}). \tag{2.33}$$

Otherwise, we can always restrict our attention on $\mathcal{N}$ to derive various approximations and then apply Lemma 2 to further approximate such results established on $\mathcal{N}$ to the counterparts on $\{1, 2, \ldots, n\}$ with negligible error term $o_p(n^{-1/2})$. Thus, without loss of generality, from now on we shall simply use $\mathcal{N} = \{1, 2, \ldots, n\}$.

Proof of Theorem 1. From (2.33), $\max_{1\le i\le n}|\hat{\mu}(X_i)-\mu(X_i)| = o_p(n^{-1/4})$; see Remark 4. Also, by Lemma 3 below, $\hat{\theta} = \theta + O_p(n^{-1/2})$. Under Assumption 3(ii), applying Taylor's expansion, we can obtain (again, as discussed in Remark 4, we take $\mathcal{N} = \{1, 2, \ldots, n\}$):

$$\hat{f}_Y(y) = \frac{1}{n}\sum_{i=1}^{n} f_e(y-\mu(X_i)|\theta) - I_n(y) + J_n(y)^T(\hat{\theta}-\theta) + O(\delta_n), \tag{2.34}$$

where

$$\begin{aligned} I_n(y) &= \frac{1}{n}\sum_{i=1}^{n} f_e'(y-\mu(X_i)|\theta)[\hat{\mu}(X_i)-\mu(X_i)], &(2.35)\\ J_n(y) &= \frac{1}{n}\sum_{i=1}^{n} \dot{f}_e(y-\mu(X_i)|\theta) \quad \text{with} \quad \dot{f}_e(z|\theta) = \frac{\partial f_e(z|\theta)}{\partial\theta}, &(2.36)\\ \delta_n &= \sup_{1\le i\le n}[\hat{\mu}(X_i)-\mu(X_i)]^2 + |\hat{\theta}-\theta|^2 = o_p(n^{-1/2}). &(2.37) \end{aligned}$$

Under Assumption 3(ii), an application of Lemma 1(ii) gives

$$\sup_{y\in\mathcal{Y}}\left|J_n(y) - \mathbb{E}[\dot{f}_e(y-\mu(X)|\theta)]\right| = o_p(1). \tag{2.38}$$

From (2.34), (2.37), (2.38) and Lemmas 3–4 below, we can obtain the uniform approximation

$$\sqrt{n}[\hat{f}_Y(y) - f_Y(y)] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\zeta_i(y) - f_Y(y)] + o_p(1), \tag{2.39}$$

uniformly over $y \in \mathcal{Y}$, where

$$\zeta_i(y) = f_e(y-\mu(X_i)|\theta) - e_i f_e'(y-\mu(X_i)|\theta) + \mathbb{E}[\dot{f}_e(y-\mu(X)|\theta)]^T \times \mathcal{I}_\theta^{-1} \times \psi(e_i|\theta) \tag{2.40}$$

From the identity (2.3), $\mathbb{E}(e_i) = 0$, and $\mathbb{E}\psi(e_i|\theta) = 0$ since $\psi(e_i|\theta)$ is the score function, the summands $\{\zeta_i(y) - f_Y(y)\}_{i=1}^n$ are iid and have zero mean. The finite-dimensional convergence easily follows from the Cramér-Wold device. Also, tightness follows since all the functions involving y has bounded derivative with respect to y (Assumption 3(ii)). Thus, $\{\sqrt{n}[\hat{f}_Y(y) - f_Y(y)], y \in \mathcal{Y}\} \Rightarrow \{G(y), y \in \mathcal{Y}\}$ for a centered Gaussian process $\{G(y)\}$ with $\mathrm{Cov}\{G(y), G(y')\} = \mathrm{Cov}\{\zeta_i(y), \zeta_i(y')\}$.

Note that, we have

$$\mathbb{E}[e_i\psi(e_i|\theta)] = \mathbb{E}\left[e_i\frac{\dot{f}_e(e_i|\theta)}{f_e(e_i|\theta)}\right] = \int_{\mathbb{R}} z\dot{f}_e(z|\theta)dz = \frac{\partial \int_{\mathbb{R}} z f_e(z|\theta)dz}{\partial\theta} = \frac{\partial\mathbb{E}(e_i)}{\partial\theta} = 0, \quad (2.41)$$

where the last "=" follows from $\mathbb{E}(e_i) = 0$ and the third "=" follows from exchanging the order of integral and derivative, which is justified via the Dominated Convergence Theorem under Assumption 3(iv). Using (2.41), $\mathbb{E}(e_i) = 0$, $\mathbb{E}\psi(e_i|\theta) = 0$, and the independence between e_i and X_i, we can verify that $\mathrm{Cov}\{\zeta_i(y), \zeta_i(y')\}$ is given by (2.16). $\diamondsuit$

Proof of Theorem 2. We only consider the uniform consistency for $\hat{\Sigma}_0(y, y')$ since the other cases $\hat{\Sigma}_\mu(y, y')$ and $\hat{\Sigma}_\theta(y, y')$ can be treated similarly. By $\hat{\theta} = \theta + O_p(n^{-1/2})$, (2.33), and Assumption 3(ii), $\sup_{1\le i\le n, y\in\mathcal{Y}} |f_e(y - \hat{\mu}(X_i)|\hat{\theta}) - f_e(y - \mu(X_i)|\theta)| = o_p(1)$. Thus,

$$\sup_{y,y'\in\mathcal{Y}}\left|\frac{1}{n}\sum_{i=1}^n f_e(y - \hat{\mu}(X_i)|\hat{\theta})f_e(y' - \hat{\mu}(X_i)|\hat{\theta}) - \frac{1}{n}\sum_{i=1}^n f_e(y - \mu(X_i)|\theta)f_e(y' - \mu(X_i)|\theta)\right| \xrightarrow{p} 0.$$

Using the same argument in the proof of (2.38), we can prove

$$\sup_{y,y'\in\mathcal{Y}}\left|\frac{1}{n}\sum_{i=1}^n f_e(y - \mu(X_i)|\theta)f_e(y' - \mu(X_i)|\theta) - \mathbb{E}[f_e(y - \mu(X)|\theta)f_e(y' - \mu(X)|\theta)]\right| \xrightarrow{p} 0.$$

Combining the above two results, we have $n^{-1}\sum_{i=1}^n f_e(y - \hat{\mu}(X_i)|\hat{\theta})f_e(y' - \hat{\mu}(X_i)|\hat{\theta}) \xrightarrow{p} \mathbb{E}[f_e(y-\mu(X)|\theta)f_e(y'-\mu(X)|\theta)]$, uniformly in $y, y' \in \mathcal{Y}$. Also, from Theorem 1, $\hat{f}_Y(y) \xrightarrow{p} f_Y(y)$, uniformly in $y \in \mathcal{Y}$. Thus, $\hat{\Sigma}_0(y, y') \xrightarrow{p} \Sigma_0(y, y')$, uniformly in $y, y' \in \mathcal{Y}$. $\diamondsuit$

Proof of Theorem 3. By Theorem 1, $\sqrt{n}[\hat{f}_Y(y) - f_Y(y)]$ converges to a Gaussian process $G(y)$ on $\mathcal{Y}$. Note that, under the newly imposed strengthened version of Assumption 3(ii), $\zeta_i(y)$ in (2.40) is k times differentiable, so the limiting process $G(y)$ is also k times differentiable. By the same argument [Equation (2.34)] in the proof of Theorem 1, we can show that $\sqrt{n}[\hat{f}_Y^{(k)}(y) - f_Y^{(k)}(y)]$ converges uniformly to a Gaussian process on $\mathcal{Y}$. Under this setting, convergence of function implies convergence of derivatives [see Rudin (1976)], and thus (2.20) holds. Since the Gaussian process $G(y)$ is k times differentiable, the derivative $G^{(k)}(y)$ is again a Gaussian process with the desired autocovariance. $\diamondsuit$

Proof of Theorem 4. Under the regularity conditions in Assumption 4, for $\hat{\beta}$ in the nonlinear least-squares estimation (2.22), we have $|\hat{\beta} - \beta| = O_p(n^{-1/2})$ and

$$\hat{\beta} - \beta = \frac{\mathcal{J}_\beta^{-1}}{n} \sum_{i=1}^{n} e_i \dot{\mu}_\beta(X_i) + o_p(n^{-1/2}). \tag{2.42}$$

See, e.g., Jennrich (1969). For $\check{f}_Y(y)$ in (2.23), by a similar Taylor expansion in (2.34),

$$\check{f}_Y(y) = \frac{1}{n} \sum_{i=1}^{n} f_e(y - \mu_\beta(X_i)|\theta) - N_n(y)^T(\hat{\beta} - \beta) + J_n(y)^T(\hat{\theta} - \theta) + o_p(n^{-1/2}), \tag{2.43}$$

where J_n is defined in (2.36) with $\mu(X_i)$ therein replaced by $\mu_\beta(X_i)$, and

$$N_n(y) = \frac{1}{n} \sum_{i=1}^{n} f_e'(y - \mu_\beta(X_i)|\theta) \dot{\mu}_\beta(X_i).$$

An application of Lemma 1(ii) gives

$$\sup_{y \in \mathcal{Y}} |N_n(y) - \mathbb{E}[f_e'(y - \mu_\beta(X)|\theta) \dot{\mu}(X; \beta)]| = o_p(1). \tag{2.44}$$

Comparing (2.34) and (2.43) and using (2.42), (2.44), and (2.50) below, we can obtain

$$\sqrt{n}[\hat{f}_Y(y) - \check{f}_Y(y)] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta_i(y) + o_p(1),$$

uniformly over $y \in \mathcal{Y}$, where

$$\eta_i(y) = \left\{ \mathbb{E}[f_e'(y - \mu_\beta(X)|\theta) \dot{\mu}_\beta(X)^T] \mathcal{J}_\beta^{-1} \dot{\mu}_\beta(X_i) - f_e'(y - \mu_\beta(X_i)|\theta) \right\} e_i.$$

As in the proof of Theorem 1, tightness and finite-dimensional convergence can be established and thus

$$\left\{ \sqrt{n}[\hat{f}_Y(y) - \check{f}_Y(y)], \quad y \in \mathcal{Y} \right\} \Rightarrow \left\{ Z(y), \quad y \in \mathcal{Y} \right\},$$

where $\{Z(y)\}$ is centered Gaussian process and $\mathrm{Cov}\{Z(y), Z(y')\} = \mathrm{Cov}\{\eta_i(y), \eta_i(y')\}$ which, after some calculations, reduces to the autocovariance function given in Theorem 4. The convergence of nT_n then follows from the continuous mapping theorem. $\diamondsuit$

Proof of Theorem 5. For model (2.27), $f_Y(y) = \phi(y; \mu, \theta)$ and $\zeta_i(y)$ in (2.40) becomes

$\zeta_i(y) = \phi(y;\mu,\theta) + e_i\dot{\phi}_\mu(y;\mu,\theta) + e_i^2\dot{\phi}_\theta(y;\mu,\theta)$. Therefore, from (2.43), we have

$$\hat{f}_Y(y) = \phi(y;\mu,\theta) + \frac{1}{n}\sum_{i=1}^{n}\Big\{e_i\dot{\phi}_\mu(y;\mu,\theta) + (e_i^2-\theta)\dot{\phi}_\theta(y;\mu,\theta)\Big\} + o_p(n^{-1/2}), \qquad (2.45)$$

uniformly over $y \in \mathcal{Y}$. On the other hand, by Taylor's expansion,

$$\hat{f}_Y^{\mathrm{MLE}}(y) = \phi(y;\mu,\theta) + (\overline{\mu}-\mu)\dot{\phi}_\mu(y;\mu,\theta) + (\overline{\theta}-\theta)\dot{\phi}_\theta(y;\mu,\theta) + O_p(n^{-1}), \qquad (2.46)$$

uniformly over $y \in \mathcal{Y}$. Using $\overline{\mu}-\mu = n^{-1}\sum_{i=1}^n e_i$ and $\overline{\theta} = n^{-1}\sum_{i=1}^n e_i^2 + O_p(1/n)$ and from (2.45)–(2.46), we conclude that $\sup_{y\in\mathcal{Y}}|\hat{f}_Y(y) - \hat{f}_Y^{\mathrm{MLE}}(y)| = o_p(n^{-1/2})$. ♢

Lemma 3. *For $\hat{\theta}$ in (2.13), under the conditions in Theorem 1, we have*

$$\hat{\theta} = \theta + \frac{\mathcal{I}_\theta^{-1}}{n}\sum_{i=1}^{n}\psi(e_i|\theta) + o_p(n^{-1/2}), \qquad (2.47)$$

where $\psi(e|\theta)$ and $\mathcal{I}_\theta$ are defined in Theorem 1. In particular, $\hat{\theta} = \theta + O_p(n^{-1/2})$.

Proof. The maximization problem (2.13) is equivalent to $\hat{\theta} = \mathrm{argmin}_\theta \frac{1}{n}\sum_{i=1}^n \log f_e(\hat{e}_i|\theta)$. Define $\hat{\theta}^* = \mathrm{argmin}_\theta \frac{1}{n}\sum_{i=1}^n \log f_e(e_i|\theta)$, which is the MLE of θ based on $e_1,\ldots,e_n$. Under regularity conditions in Assumption 3(iv), the classical result in MLE states that $\hat{\theta}^*$ satisfies the asymptotic representation (2.47). Thus, from the classical proof of consistency and asymptotic normality of MLE, to show that $\hat{\theta}$ also satisfies (2.47), it suffices to establish the uniform convergence

$$\Delta_n \;:=\; \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\log f_e(\hat{e}_i|\theta) - \frac{1}{n}\sum_{i=1}^{n}\log f_e(e_i|\theta)\right| \xrightarrow{p} 0. \qquad (2.48)$$

From (2.33),

$$\max_{1\le i\le n}|\hat{e}_i - e_i| = \max_{1\le i\le n}|\hat{\mu}(X_i) - \mu(X_i)| \xrightarrow{p} 0. \qquad (2.49)$$

Thus, for $\epsilon > 0$ in Assumption 3(iii), for large enough n, $\max_{1\le i\le n}|\hat{e}_i - e_i| < \epsilon$ with probability tending to one. By mean value theorem and Assumption 3(iii),

$$\Delta_n \;\le\; \frac{1}{n}\sum_{i=1}^{n}\sup_{\theta\in\Theta}|\log f_e(\hat{e}_i|\theta) - \log f_e(e_i|\theta)|$$

$$\leq \frac{1}{n}\sum_{i=1}^{n}|\hat{e}_i - e_i|\ell_1(e_i) \qquad \Big(\ell_1(e_i) \text{ is defined in Assumption 3(iii)}\Big)$$
$$\leq \max_{1\leq i\leq n}|\hat{e}_i - e_i| \times \frac{1}{n}\sum_{i=1}^{n}\ell_1(e_i) \xrightarrow{p} 0,$$

in view of (2.49) and $n^{-1}\sum_{i=1}^{n}\ell_1(e_i) \xrightarrow{p} \mathbb{E}\ell_1(e) < \infty$. This completes the proof. ♢

Lemma 4. *For $I_n(y)$ in (2.35), under conditions in Theorem 1, we have*

$$I_n(y) = \frac{1}{n}\sum_{j=1}^{n} e_j f'_e(y - \mu(X_j)|\theta) + o_p(n^{-1/2}), \quad \textit{uniformly over } y \in \mathcal{Y}. \tag{2.50}$$

Proof. For convenience, we write $K_{h_n}(u) = K(u/h_n)/h_n$. Rewrite

$$\hat{\mu}(x) = \sum_{i=1}^{n} w_i(x)Y_i \quad \text{with} \quad w_i(x) = \frac{K_{h_n}(x - X_i)}{n\hat{f}_X(x)}, \quad \hat{f}_X(x) = \frac{1}{n}\sum_{i=1}^{n} K_{h_n}(x - X_i). \tag{2.51}$$

Using $\sum_{i=1}^{n} w_i(x) = 1$ and $Y_i = \mu(X_i) + e_i$, we can obtain

$$\hat{\mu}(x) - \mu(x) = B_n(x) + \sum_{i=1}^{n} w_i(x)e_i \quad \text{with} \quad B_n(x) = \sum_{i=1}^{n} w_i(x)[\mu(X_i) - \mu(x)]. \tag{2.52}$$

The bias term $B_n(x)$ satisfies $\sup_x |B_n(x)| = O_p(h_n^2)$. Thus, by (2.52) and the uniform boundedness of $f'_e(z|\theta)$ (Assumption 3(ii)), we have

$$I_n(y) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} w_j(X_i)e_j f'_e(y - \mu(X_i)|\theta) + O_p(h_n^2). \tag{2.53}$$

For $\hat{f}_X(x)$ in (2.51), by property of nonparametric kernel density estimator,

$$\hat{f}_X(X_i) = f_X(X_i) + O_p\Big(h_n^2 + \sqrt{\frac{\log n}{nh_n}}\Big), \quad \text{uniformly in } i = 1, 2, \ldots. \tag{2.54}$$

Again, as discussed in Remark 4, here we have used all $i \in \{1, 2, \ldots, n\}$ instead of $i \in \mathcal{N}$. Substituting this into $w_i(x)$ in (2.51), by (2.53) we have

$$I_n(y) = \left[1 + O_p\Big(h_n^2 + \sqrt{\frac{\log n}{nh_n}}\Big)\right]\tilde{I}_n(y) + O_p(h_n^2), \tag{2.55}$$

where

$$\tilde{I}_n(y) = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}\xi_{ij}(y)e_j \quad \text{with} \quad \xi_{ij}(y) = \frac{K_{h_n}(X_i - X_j)f'_e(y-\mu(X_i)|\theta)}{f_X(X_i)}.$$

Further write

$$\begin{aligned}\tilde{I}_n(y) &= \frac{1}{n^2}\sum_{i=1}^{n}\xi_{ii}(y)e_i + \frac{1}{n^2}\sum_{i\neq j}\{\xi_{ij}(y) - \mathbb{E}[\xi_{ij}(y)|X_j]\}e_j + \frac{1}{n^2}\sum_{i\neq j}e_j\mathbb{E}[\xi_{ij}(y)|X_j] \\ &:= \tilde{I}_{n,0}(y) + \tilde{I}_{n,1}(y) + \tilde{I}_{n,2}(y). \qquad (2.56)\end{aligned}$$

For $\tilde{I}_{n,0}(y)$, by Lemma 1(ii), we have

$$\tilde{I}_{n,0}(y) = \frac{K(0)}{n^2 h_n}\sum_{i=1}^{n}\frac{f'_e(y-\mu(X_i)|\theta)}{f_X(X_i)} = O_p\Big(\frac{1}{nh_n}\Big) = o_p(n^{-1/2}),$$

uniformly over $y \in \mathcal{Y}$. Therefore, from (2.56), in order to prove (2.50), it suffices to establish:

$$\sup_{y\in\mathcal{Y}}|\tilde{I}_{n,1}(y)| = o_p(n^{-1/2}) \text{ and } \sup_{y\in\mathcal{Y}}\left|\tilde{I}_{n,2}(y) - \frac{1}{n}\sum_{j=1}^{n}e_j f'_e(y-\mu(X_j)|\theta)\right| = o_p(n^{-1/2}). \qquad (2.57)$$

[proof of the first assertion in (2.57)]: Essentially, (2.56) is a Hoeffding-type decomposition, and similar to the property of Hoeffding decomposition, it can be verified that $\{\xi_{ij}(y) - \mathbb{E}[\xi_{ij}(y)|X_j]\}e_j, 1 \le i \neq j \le n$, are uncorrelated so

$$\|\tilde{I}_{n,1}(y)\|^2 = \frac{1}{n^4}\sum_{i\neq j}\|\{\xi_{ij}(y) - \mathbb{E}[\xi_{ij}(y)|X_j]\}e_j\|^2. \qquad (2.58)$$

Note that we have the following facts:

(i) $\xi_{ij}(y) - \mathbb{E}[\xi_{ij}(y)|X_j]$ and e_j are independent.

(ii) $\|\xi_{ij}(y) - \mathbb{E}[\xi_{ij}(y)|X_j]\| \le \|\xi_{ij}(y)\| + \|\mathbb{E}[\xi_{ij}(y)|X_j]\| \le 2\|\xi_{ij}(y)\|$ via the triangle inequality and Jensen's inequality $\|\mathbb{E}[\xi_{ij}(y)|X_j]\| \le \|\xi_{ij}(y)\|$.

(iii) $|\xi_{ij}(y)| \le c_1 K_{h_n}(X_i - X_j)$ for some constant c_2, in view of the boundedness of $f'_e(z|\theta)$ and $f_X(x) > c$ in Assumption 2(iv).

(iv) $\mathbb{E}K^2_{h_n}(X_i - X_j) = O(1/h_n)$ for $i \neq j$.

Combining the above facts with (2.58), we see that there exists constant c_3 such that

$$\|\tilde{I}_{n,1}(y)\|^2 \leq \frac{c_3}{n^4}\frac{n^2-n}{h_n} = o(n^{-1}), \quad \text{for all } y \in \mathcal{Y}. \tag{2.59}$$

For the derivative $\tilde{I}'_{n,1}(y) = n^{-2}\sum_{j=1}^n \sum_{i=1}^n \{\xi'_{ij}(y) - \mathbb{E}[\xi'_{ij}(y)|X_j]\}e_j$, the same argument in (2.58)–(2.59) yields $\|\tilde{I}'_{n,1}(y)\|^2 = o(1/n)$. By Lemma 1(ii), the first assertion in (2.57) follows.

[proof of the second assertion in (2.57)]: For $\tilde{I}_{n,2}(y)$, we calculate

$$\begin{aligned} \mathbb{E}[\xi_{ij}(y)|X_j] &= \int_{-\infty}^{\infty} \frac{1}{h_n} K\left(\frac{x-X_j}{h_n}\right) f'_e(y-\mu(X_i)|\theta)dx \\ \left(\text{by transformation } u = \frac{x-X_j}{h_n}\right) &= \int_{-\infty}^{\infty} K(u) f'_e(y-\mu(X_j+uh_n)|\theta)du. \end{aligned} \tag{2.60}$$

In (2.60), applying Taylor's expansion $f'_e(y-\mu(X_j+uh_n)|\theta) = f'_e(y-\mu(X_j)|\theta) + f''_e(y-\mu(X_j)|\theta)\mu'(X_j)uh_n + O(u^2h_n^2)$, where all the derivatives here are uniformly bounded, and using the conditions in Assumption 2(i), we can obtain

$$\sup_{y\in\mathcal{Y}} |\mathbb{E}[\xi_{ij}(y)|X_j] - f'_e(y-\mu(X_j)|\theta)| = O(h_n^2), \quad \text{for all } i,j. \tag{2.61}$$

The second assertion in (2.57) follows from (2.61) under conditions $nh_n^4 \to 0$, $\mathbb{E}|e_j| < \infty$. $\diamondsuit$

Chapter 3
Stationary Density Estimation for ARCH Models with Applications in Quantile Regression

3.1 Introduction

The autoregressive conditional heteroscedastic (ARCH) models proposed by Engle (1982) are helpful tools for analyzing the variance of time series data. The fundamental idea of ARCH models is that the conditional variances are non-constant but change over time. Thus, the application of these models is widespread in situations where the goal is to analyze and forecast volatility since this volatility appears to change over time and exhibit clustering, which are periods of swings interspersed with periods of relative calm [Bollerslev et al. (1994)]. For instance, in finance, one objective is to model the asset returns volatility, as this volatility is considered a measure of risk, and investors want a premium for investing in risky assets [Teräsvirta and Zhao (2011)]. The observations of these asset returns appear to be uncorrelated or nearly uncorrelated, but the data can contain higher-order dependence. Thus, to parameterize this type of dependence, the ARCH model is the most popular model used; see Engle (2001) for examples of applications.

Let Z_i be a real-valued stationary time series of interest, $1 \leq i \leq n$, from the nonparametric autoregressive conditional heteroscedasticity (NARCH) model of the form

$$Z_i = \sigma\left(Z_{i-1}\right) e_i \quad \text{with} \quad e_i \overset{\text{i.i.d.}}{\sim} N(0,1), \tag{3.1}$$

where $\sigma(\cdot) > 0$ is the unknown conditional variance (or volatility) function, and e_i are i.i.d. error terms which are independent of the past series values Z_{i-1}, Z_{i-2}, ... The nonparametric model (3.1) does not impose a specific structure on $\sigma(\cdot)$, which lets the data speak for themselves and, in some cases, avoid mis-specification. Also, the normality assumption of the error term is one of the most common assumptions in nonlinear regression, time series, and discrete samplings from diffusion models, and it can be satisfied for many practical data after proper transformations. This assumption is not as restrictive as imposing a parametric form for $\sigma(\cdot)$, which can be difficult or even unrealistic to know in advance in real data. For example, in financial econometrics applications, for discretely sampled data from stochastic diffusion models driven by a standard Brownian motion, the error term is normally distributed[1]. These discrete models are a kind of application of the model (3.1).

The model (3.1) can generate heavy-tailed distributions as *kurtosis*$(Z_i) \geq 3$; see Zhao (2008) for details. Thus, the heavy-tailedness characteristic of this model makes it a helpful candidate in many financial and economic applications, where it is frequently observed that the time series exhibit heavy tails. Due to the so-called *curse of dimensionality*, it is practically infeasible to extend the nonparametric model (3.1) to orders beyond two. We observe that model (3.1) is a generalization of the ARCH model of order 1 in Engle (1982), which is of the form $Z_i = \sqrt{\alpha_0 + \alpha_1 Z_{i-1}^2}e_i$, where α_0 and α_1 are unknown real parameters. Other extension of model (3.1) include the ARCH model of order p.

The marginal density function plays an important role in the statistical analysis of time-dependent data, as it contains all distributional information about the data. For example, in applications in economic and financial time series [Silverman (1986); Liao and Stachurski (2015)], the marginal density estimation often provides valuable information about the shape and tail behavior of the distribution of market indices, asset returns, interest rates, GDP growth, inflation, and so on, which is a real need to capture the large fluctuations of these time series. Nonparametric density estimation has been extensively investigated in the literature; in particular, the nonparametric kernel density estimator, the so-called Parzen–Rosenblatt, is one of the most widely used methods for this problem.

[1]For a continuous-time diffusion model: $\mathrm{d}Y_t = \sigma(Y_t)\mathrm{d}B_t$, for $t \geq 0$ and $\{B_t\}_{t\geq 0}$ a standard Brownian motion, we have in practice for discrete observations, sample at discrete interval of length $\Delta > 0$ and $t = i\Delta$, we can approximate this model by: $Y_{(i+1)\Delta} - Y_{i\Delta} = \sigma(Y_{i\Delta})\Delta^{1/2}\Delta^{-1/2}[B_{(i+1)\Delta} - B_{i\Delta}]$. Write $Z_i = Y_{(i+1)\Delta} - Y_{i\Delta}$, $Y_i = Y_{i\Delta}$ and $e_i = \Delta^{-1/2}[B_{(i+1)\Delta} - B_{i\Delta}]$, we have that $Z_i = \sigma(Y_i)\Delta^{1/2}e_i$. Then according to the properties of the Brownian motion, we have that e_i has a standard normal distribution.

The properties of this estimator have been studied by Robinson (1983), Ango-Nze and Portier (1994), Ango-Nze and Doukhan (1998), Doukhan and Louhichi (2001) or Roussas (2000) under various types of mixing conditions, by Bosq (1998) in the context of strong mixing sequences, by Dedecker et al. (2007) for different scenarios of weak dependence, and, by Hansen (2008) for stationary strong mixing multivariate data. Denote by $f_Z(z)$ the density of Z. Given $Z_1, \ldots, Z_n$, the Parzen–Rosenblatt estimator [See Silverman (1986)] of $f_Z(z)$ is

$$\tilde{f}_Z(z) = \frac{1}{nb_n} \sum_{i=1}^{n} K\left(\frac{z - Z_i}{b_n}\right), \tag{3.2}$$

for kernel function $K(\cdot)$ (an integrable function that integrates to 1) and bandwidth $b_n \to 0$. The estimator presented in equation (3.2) does not require knowledge of the density form of the data. This feature has made it a widely used choice in various fields, including economics and finance, where the underlying density of interest is often unknown, particularly at the beginning of a study.

Under appropriate conditions [See Li and Racine (2007)], $\tilde{f}_Z(z)$ is $\sqrt{nb_n}$-consistent. This slow convergence rate $\sqrt{nb_n}$ of the classical nonparametric kernel density estimator (3.2) can be problematic for small sample sizes. In addition, the density estimator in (3.2) does not depend on the subjacent model of the process in (3.1), being robust to the model assumptions. However, by not considering that particular structure of the process; hence, this estimator is not the most efficient one.

To overcome the aforementioned slow convergence issue of the estimator in (3.2), $\sqrt{n}$-consistent marginal density estimates have been proposed by Saavedra and Cao (1999, 2000), Schick and Wefelmeyer (2004) for moving average processes, by Kim and Wu (2007) for nonlinear autoregressive models of order one with constant variance, by Schick and Wefelmeyer (2007) for invertible linear processes, by Zhao (2010) for nonlinear autoregressive conditional heteroscedastic models with parametric noise, in which the structure for the functions of mean and variance is known, by Kim et al. (2015) for nonlinear and homocedastic autoregressive processes with nonparametric noise for which the conditional mean has a parametric specification, by Li and Tu (2016) for semiparametric regression models where the mean function is specified as a parametric form up to unknown parameters and the error density is left unspecified, by Delaigle et al. (2016) for the log-volatility of generalized autoregressive conditionally heteroscedastic

GARCH(1,1) processes, and, more recently, by Truquet (2019) for autoregressive time series models for which the conditional mean and variance have a parametric specification.

This Chapter has two main goals. For our first goal, motivated by the fact that nonparametric density estimation is a relevant tool for the analysis of time series, as it contains all the distributional information about the data, we propose an $\sqrt{n}$-consistent density estimate for $f_Z(z)$, the marginal density function of Z_i, by utilizing the specific structure of (3.1). Instead of imposing parametric form on $\sigma(\cdot)$ as is often used in traditional approaches, we impose the normality assumption on noise e while leaving $\sigma(\cdot)$ completely unspecified. In addition to the faster convergence rate over the nonparametric kernel density estimator (3.2), the proposed estimator has several other appealing properties. For example, it is uniformly bounded, and possesses certain smoothness properties which the classical nonparametric kernel density estimator lacks. The second goal is to get a more accurate estimate of the limiting variance of the estimated coefficients in a quantile regression model, whose errors follow the nonparametric ARCH structure in (3.1). We do this by estimating the density function of the error terms in the model $f_Z(z)$, using the proposed efficient estimation of the marginal density.

The rest of this Chapther is structured as follows. Section 3.2 presents the main results: $\sqrt{n}$-consistency of our proposed density estimator is shown in Subsection 3.2.1, and an application of the proposed density estimator to quantile regression with nonparametric autoregressive conditional heteroscedastic errors is described in Subsection 3.2.2. In Section 3.3, we examine the performance of our density estimate through a simulation study (see Subsection 3.3.1) and then apply our method to the S&P 500 Index data (refer to Subsection 3.3.2). Also, we present both simulation data (outlined in Subsection 3.3.3) and an empirical analysis (see Subsection 3.3.4) to observe the proposed methodology's performance in estimating the limiting variance of the estimated coefficients in quantile regression with nonparametric ARCH errors. Proofs are gathered in Section 3.4.

3.2 Main results

Throughout, denote by $\phi(z;\sigma^2)$ the normal density of $N(0,\sigma^2)$ and by $\dot{\phi}(z;\sigma^2)$ the derivative of $\phi(z;\sigma^2)$ with respect to σ^2.

3.2.1 Stationary density estimation for nonparametric ARCH model

We propose an efficient $\sqrt{n}$-consistent density estimate for f_Z by utilizing the specific structure of (3.1). In model (3.1), conditioning on Z_{i-1}, we have that Z_i has the density function $\phi\left(z;\sigma^2\left(Z_{i-1}\right)\right)$. Thus, by conditioning argument of density function,

$$f_Z(z) = \mathbb{E}\phi\left(z;\sigma^2\left(Z_{i-1}\right)\right). \tag{3.3}$$

If $\sigma^2(\cdot)$ were known, using the sample version of (3.3), we can estimate $f_Z(z)$ by

$$f_Z^*(z) = \frac{1}{n}\sum_{i=1}^{n}\phi\left(z;\sigma^2\left(Z_{i-1}\right)\right). \tag{3.4}$$

When $\sigma^2(\cdot)$ is unknown, it is necessary to estimate it. Let $\hat{\sigma}^2(\cdot)$ be the nonparametric estimate of $\sigma^2(\cdot)$. In view of (3.4), we propose estimating the stationary density $f_Z(z)$ by

$$\hat{f}_Z(z) = \frac{1}{n}\sum_{i=1}^{n}\phi\left(z;\hat{\sigma}^2\left(Z_{i-1}\right)\right). \tag{3.5}$$

For $\hat{\sigma}^2(\cdot)$ in (3.5), we use the widely used nonparametric kernel smoothing estimator (Nadaraya-Watson kernel-weighted) given by

$$\hat{\sigma}^2\left(z\right) = \frac{\sum_{i=1}^{n} Z_i^2 K\left(\left(z - Z_{i-1}\right)/h_n\right)}{\sum_{i=1}^{n} K\left(\left(z - Z_{i-1}\right)/h_n\right)}, \tag{3.6}$$

for proper kernel function $K(\cdot)$ and bandwidth $h_n \to 0$. Due to the dependence structure in time series data, the optimal bandwidth for dependent data differs from the bandwidth under independence [Herrmann et al. (1992)]. For dependent data, the optimal bandwidth can be expressed as $h_n = \nu h_n^*$, where h_n^* is the bandwidth calculated as if the data were independent [Ruppert et al. (1995)], and $\nu > 1$ is a variance correction factor due to dependence structure. However, estimating ν is generally challenging [Altman (1990);Wu and Zhao (2007)]. For example, based on simulation studies, Li and Zhao (2019) recommend using $\nu = 2$ for their dependent data structure. Our simulations showed that using a constant $\nu = 2$ and the automatic plug-in bandwidth selector h_n^* of Ruppert et al. (1995) (implemented using `R` function `dpill` in the `R` package `KernSmooth`) resulted in a better performance for our proposed estimator in terms of mean squared error. Therefore, we set $h_n = 2h_n^*$ as our approach for selecting the bandwidth h_n in (3.6) for the simulations and applications discussed in Section 3.3.

Now, we establish the interesting result that the estimate of $f_Z(z)$ in (3.5) can achieve $\sqrt{n}$ parametric convergence rate, despite the use of the nonparametric estimate $\hat{\sigma}^2(\cdot)$ in (3.6). Additionally to the faster convergence rate over the nonparametric kernel density estimator $\tilde{f}_Z(z)$ in (3.2), the proposed estimator $\hat{f}_Z(z)$ has other appealing properties, such as being uniformly bounded and possessing smoothness features which the classical nonparametric kernel density estimator $\tilde{f}_Z(z)$ lacks for small sample sizes. We now give some assumptions for establishing the $\sqrt{n}$-consistency of the proposed density estimate of $f_Z(z)$ in (3.5).

Assumption 5. (i) The kernel function $K(\cdot)$ is symmetric and continuously differentiable, has bounded derivative and bounded support, satisfies $\int_{\mathbb{R}} K(u)du = 1$ and $\int_{\mathbb{R}} uK(u)du = 0$ (implication of symmetry). (ii) h_n satisfies $nh_n^4 \to 0$, $\sqrt{n}h_n/(\log n)^2 \to \infty$. (iii) $\sigma(\cdot)$ is twice continuously differentiable. There exists $\sigma_0 > 0$ such that $\sigma(z) \geq \sigma_0$ for all $z \in \mathbb{R}$.

Assumption 6. (i) $\{e_i\}$ are i.i.d. and $\{Z_i\}$ is causal in the sense that Z_i depends only on $\mathcal{F}_i$, the sigma field generated by $\{e_j\}_{j \leq i}$. (ii) The sequence $\{Z_i\}$ is strictly stationary and strongly mixing with $\alpha-$mixing coefficients $\alpha_k = O(\rho^k)$ for some $\rho \in (0, 1)$. Furthermore, $\mathbb{E}(|Z_i|^s) < \infty$ for some $s > 2$.

Condition (i) in Assumption 5 is typical in nonparametric kernel smoothing literature. We impose boundedness on the support of $K(\cdot)$ for the brevity of proofs; it may be removed at the cost of lengthier proofs. In particular, the Gaussian kernel is allowed because its tails are thin. In nonparametric regression, the usual bandwidth conditions are $h_n \to 0$ and $nh_n \to \infty$, and condition (ii) in Assumption 5 imposes a stronger bandwidth condition in order to suppress the bias effect and establish $\sqrt{n}$ convergence rate of our estimator. Assumption 6(ii) is frequently used to control dependence structures in time series, and it is reasonably weak and fulfilled for many time series models. Also, the exponentially decaying mixing coefficient assumption can be replaced by some polynomial decaying assumption with much lengthier proofs; See Hansen (2008).

The following Theorem describes the asymptotic behavior of our proposed estimator $\hat{f}_Z(z)$.

Theorem 6. *For $\hat{f}_Z(z)$ in (3.5) with $\hat{\sigma}(\cdot)$ defined in (3.6), under Assumptions 5 and 6, for any bounded interval $\mathcal{Z}$, the functional Central Limit Theorem (CLT) holds:*

$$\left\{\sqrt{n}\left[\hat{f}_Z(z) - f_Z(z)\right], \quad z \in \mathcal{Z}\right\} \Rightarrow \{G(z), \quad z \in \mathcal{Z}\} \tag{3.7}$$

for a centered Gaussian process $\{G(z)\}$ *with autocovariance* $\Sigma(z, z') = \mathrm{Cov}\{G(z), G(z')\}$ *given by*

$$\Sigma(z, z') = \gamma_0(z, z') + \sum_{k=1}^{\infty}[\gamma_k(z, z') + \gamma_k(z', z)] \quad \textit{with} \quad \gamma_k(z, z') = \mathbb{E}[\xi_i(z)\xi_{i+k}(z')], \quad (3.8)$$

where

$$\xi_i(z) = [\phi(z; \sigma^2(Z_{i-1})) - \mathbb{E}\phi(z; \sigma^2(Z_{i-1}))] + \dot{\phi}\left(z; \sigma^2(Z_{i-1})\right)\sigma^2(Z_{i-1})(e_i^2 - 1). \quad (3.9)$$

Here we give an intuitive explanation for the covariance structure in (3.8). For time series data $X_1, \ldots, X_n$, the CLT for the sample mean $\bar{X}$ states that, under appropriate dependence conditions, $\sqrt{n}[\bar{X} - \mathbb{E}(X_0)] \Rightarrow N(0, \sigma^2)$, where $\sigma^2 = \mathrm{Cov}(X_0, X_0) + 2\sum_{k=1}^{\infty}\mathrm{Cov}(X_0, X_k)$ is the long-run variance. In (3.8), $\Sigma(z, z')$ has a similar long-run variance structure. Note that, since Z_{i-1} and e_i are independent (the causal condition in Assumption 6), $\mathbb{E}[\dot{\phi}(z; \sigma^2(Z_{i-1}))\sigma^2(Z_{i-1})(e_i^2 - 1)] = 0$ and $\mathbb{E}\xi_i(z) = 0$. Therefore, in the absence of dependence, $\gamma_i(z, z') = 0$ for all $i \geq 1$, and $\Sigma(z, z')$ would contain only the first term $\gamma_0(z, z')$.

In (3.9), we observe that $\xi_i(z)$ has two components $\xi_{i,1}(z) := [\phi(z; \sigma^2(Z_{i-1})) - \mathbb{E}\phi(z; \sigma^2(Z_{i-1}))]$ and $\xi_{i,2}(z) := \dot{\phi}(z; \sigma^2(Z_{i-1}))\sigma^2(Z_{i-1})(e_i^2 - 1)$. The first component $\xi_{i,1}(z)$ results from the variability introduced by using the sample version (3.4) to estimate the population version (3.3), and $f_Z^*(z)$ in (3.4) admits the same function CLT in Theorem 6 with $\xi_i(z)$ therein being replaced by just the first component $\xi_{i,1}(z)$. As can be seen from the proof of Theorem 6, the second component $\xi_{i,2}(z)$ stems from the variability caused by the estimate $\hat{\sigma}^2(z)$ of $\sigma^2(z)$.

By Theorem 6, for each fixed z, we have that

$$\sqrt{n}[\hat{f}_Z(z) - f_Z(z)] \Rightarrow N\Big(0, \Sigma(z, z')\Big). \quad (3.10)$$

We observe that the proposed estimate $\hat{f}_Z(z)$ in (3.5) posses a faster convergence rate $\sqrt{n}$ compared to convergence rate $\sqrt{nb_n}$ of $\tilde{f}_Z(z)$ in (3.2). Intuitively, the estimate $\hat{f}_Z(z)$ makes use of the information from the conditional heteroscedasticity structure which, while being nonparametric, can still provide information to the density through the identity (3.3), which leads to more efficient density estimation. Below, we highlight some additional appealing characteristics of this new density estimator.

(a) *Uniform boundedness.* Since $\phi\left(z;\sigma^2(z_{i-1})\right)$ is uniformly bounded, $\hat{f}_Z(z)$ has the same uniform bound. In statistical inference, we often require the density function be bounded, thus $\hat{f}_Z(z)$ satisfies this property. However, for $\tilde{f}_Z(z)$ in (3.2), even though $K(\cdot)$ is bounded, there always exist sample paths (with asymptotically vanishing probability though) such that all Z_i 's are close to z so that $\tilde{f}_Z(z) = O\left(1/b_n\right)$ can be arbitrarily large as $b_n \to 0$.

(b) *Smoothness of the limiting process* $\{G(z)\}$ *in Theorem 6.* Since $\xi_i(z)$ in (3.9) is differentiable, the process $\{G(z)\}$ is differentiable. In general, the limiting process $\{G(z)\}$ is infinitely differentiable. However, the nonparametric kernel density estimator $\tilde{f}_Z(z)$ in (3.2) exhibits completely different behavior. To see this, let

$$U_n(z) = \frac{\sqrt{nb_n}[\tilde{f}_Z(z) - f_Z(z) - b_n^2 f_Z''(z)\int_{\mathbb{R}} u^2 K(u)du/2]}{\sqrt{f_Z(z)\int_{\mathbb{R}} K^2(u)du}}.$$

Under Assumption 5, it is well known that $U_n(z) \Rightarrow N(0,1)$ for each fixed z and that $U_n(z)$ and $U_n(z')$ are asymptotically independent $N(0,1)$ for $|z - z'| > 2Lb_n$, where we assume $K(\cdot)$ has bounded support $[-L, L]$. If we take $|z - z'| = 3Lb_n \to 0$, then $U_n(z) - U_n(z') \Rightarrow N(0,2)$. That is, even when z and z' are arbitrarily close, the distance $U_n(z) - U_n(z')$ is a non-degenerate random variable $N(0,2)$, indicating discontinuity. Roughly speaking, this means that, for large enough n, $U_n(z)$ consists of pure jumps without any continuity.

(c) *Functional CLT.* The undesirable non-smoothness of $\tilde{f}_Z(z)$ discussed in (b) is caused by the lack of tightness, which is why $\tilde{f}_Z(z)$ only lets pointwise asymptotic normality and does not admit the type of functional CLT of $\hat{f}_Z(z)$ as in Theorem 6.

To apply Theorem 6 for statistical inference purposes (for instance, confidence interval construction), now we discuss the estimation of the covariance function $\Sigma(z, z')$ in (3.8) and we will do so by estimating $\gamma_i(z, z')$. A natural estimate of $\gamma_k(z, z')$ is to replace the population version $\mathbb{E}[\xi_i(z)\xi_{i+k}(z')]$ by its sample version:

$$\hat{\gamma}_k(z, z') = \frac{1}{n-k}\sum_{i=1}^{n-k} \hat{\xi}_i(z)\hat{\xi}_{i+k}(z'), \tag{3.11}$$

where $\hat{\xi}_i(z)$ is an estimate of $\xi_i(z)$ given by

$$\hat{\xi}_i(z) = [\phi(z; \hat{\sigma}^2(Z_{i-1})) - \hat{f}_Z(z)] + \dot{\phi}\left(z; \hat{\sigma}^2\left(Z_{i-1}\right)\right)[Z_i^2 - \hat{\sigma}^2(Z_{i-1})]. \tag{3.12}$$

Note that, in (3.9), $\sigma^2(Z_{i-1})(e_i^2 - 1) = Z_i^2 - \sigma^2(Z_{i-1})$, which leads us to use $Z_i^2 - \hat{\sigma}^2(Z_{i-1})$ in (3.12).

Proposition 1. *Under the same conditions in Theorem 6, we have uniform consistency*

$$\sup_{z,z' \in \mathcal{Z}} |\hat{\gamma}_k(z, z') - \gamma_k(z, z')| = o_p(1).$$

With the estimate $\hat{\gamma}_k(z, z')$, from (3.8), we can estimate $\Sigma(z, z')$ by

$$\hat{\Sigma}(z, z') = \hat{\gamma}_0(z, z') + \sum_{k=1}^{L} [\hat{\gamma}_k(z, z') + \hat{\gamma}_k(z', z)], \tag{3.13}$$

where L is the lag truncation parameter. For long run variance estimation under time series setting, various approaches have been proposed [Hansen (1992), White (1984), Newey and West (1987) and Andrews (1991)], and (3.13) is a special case with truncated kernel. As discussed in the aforementioned works, the optimal choice for lag truncation parameter $L = Cn^{1/3}$, where C depends on other unknown quantities that can be difficult to estimate. In practice, an ad hoc choice is $L = \lfloor n^{1/3} \rfloor$.

According to Chapter 1, the classical nonparametric kernel estimator $\tilde{f}_Z(z)$ possesses bias. Thus, to conclude this section, we briefly discuss the issue of bias in estimating the density $f_Z(z)$ using $\tilde{f}_Z(z)$ in (3.2). From Li and Racine (2007), we have that the bias of $\tilde{f}_Z(z)$, at any point z, can be calculated as follows:

$$\text{Bias}\left(\tilde{f}_Z(z)\right) = [1 + o(1)]\frac{b_n^2 f_Z''(z)}{2}\int_{\mathbb{R}} u^2 K(u) du, \tag{3.14}$$

where the expression $f_Z''(z) = \partial^2 f_Z(z)/\partial z^2$ represents the second-order derivative of density $f_Z(z)$. As explained by Racine (2008), the above equation for bias has the pointwise property, which means it holds at any point z. We can observe the following facts from (3.14). First, the bias of $\tilde{f}_Z$ depends on the bandwidth b_n. As b_n decreases, the bias falls. Second, this bias is caused by the curvature of the density function, which means that the bias may increase with $f_Z''(z)$, hence is highest in the peaks of the distribution. This is because, at the peaks of the distribution, $\tilde{f}_Z$ tends to smooth the

distribution too much, thereby making the density function estimation smoother (less curved) than it actually is (we will observe this behavior through a simulation study in Subsection 3.3.3). However, as long as the bandwidth $b_n \to 0$ as $n \to \infty$ (bias $\to 0$), then the bias associated with $f''_Z(z)$ will become smaller as the available data increase, and eventually disappear. For further details, please refer to Racine (2008) or Li and Racine (2007).

Now, the bias of the classical nonparametric kernel density estimator in (3.14), under the model in (3.1) is described in the following Corollary 2. Additionally, in Subsection 3.3.3, we present a simulation study that aims to quantify this bias of $\tilde{f}_Z(z)$.

Corollary 2. *For the bias of $\tilde{f}_Z(z)$ in (3.14), with Gaussian kernel function $K(\cdot)$ and bandwidth $b_n \to 0$, if we consider the nonparametric autoregressive conditional heteroscedasticity (NARCH) model in (3.1), then the bias of $\tilde{f}_Z(z)$ is of the form of* $\text{Bias}(\tilde{f}_Z(z)) = [1 + o(1)]b_n^2 C_Z(z)$, *where*

$$C_Z(z) = \frac{1}{2}\mathbb{E}\left[\frac{1}{\sigma^3(Z)}\left(\frac{z^2}{\sigma^2(Z)} - 1\right)\phi\left(\frac{z}{\sigma(Z)}; 1\right)\right]. \tag{3.15}$$

From (3.15) in Corollary 2, we can observe two particular cases. Firstly, if $z = 0$, then

$$C_Z(0) = -\frac{1}{2\sqrt{2\pi e}}\mathbb{E}\left[\frac{1}{\sigma^3(Z)}\right] < 0.$$

Thus, in this scenario, the estimator $\tilde{f}_Z$ will tend to underestimate the true density f_Z, and this issue can be significant when $\sigma(z)$ is relatively small. Secondly, if z is much larger than the typical values of $\sigma(Z)$, then $z^2/\sigma^2(Z) - 1 > 0$ and $C_Z(z) > 0$. For example, if $c^* < \sigma(Z) < C^*$ for some $0 < c^* < C^*$, then $C_Z(z) > 0$ for all $z > C^*$. In particular, if $z \approx \pm\infty$, then $C_Z(z) > 0$. In this scenario, the estimator $\tilde{f}_Z$ will show a tendency to overestimate the true density f_Z for large positive or negative values of z, i.e., overestimate the extreme tails. These features of bias for the classical nonparametric kernel density estimator will help us to explain the performance of this estimator, compared with our proposed estimator $\hat{f}_Z(z)$, in estimating the limiting variance of the estimated coefficients in the quantile regression model with nonparametric ARCH errors in the simulation study in Subsection 3.3.3.

3.2.2 Quantile regression with nonparametric ARCH errors

Quantile regression has recently received special attention in studying time series to address important issues in many disciplines, such as economics and finance. This method introduced by Koenker and Bassett (1978) provides a way of estimating the conditional quantile function and, thus, the whole conditional distribution of the variable of interest on the explanatory variables in the regression model. It also broadens the modeling options available for time series analysis. Quantile regression has been applied to diverse statistical models as it is robust against the presence of outliers and allows extracting helpful information to make further inferences. For a general review of quantile regression, see the monograph by Koenker (2005) or Koenker et al. (2018) which also shows a wide range of additional examples. This section aims to obtain a more accurate estimate of the limiting variance of estimated coefficients in quantile regression with nonparametric ARCH errors. We achieve this by efficiently estimating the marginal density of the error term using the proposed density estimator $\hat{f}_Z(\cdot)$.

Consider observations Y_i, $1 \leq i \leq n$, from the linear regression model with nonparametric ARCH errors

$$Y_i = X_i^T \beta + Z_i \quad \text{with} \quad Z_i \sim \text{ARCH model in (3.1)}, \tag{3.16}$$

where $\beta = (\beta_0, \beta_1, \ldots, \beta_{d-1})^T \in \mathbb{R}^{d\times 1}$ is a d-dimensional coefficient vector of the vector of regressors $X_i = (1, X_{1i}, \ldots, X_{(d-1)i}) \in \mathbb{R}^{1\times d}$ in the regression model and the error term Z_i is independent of X_j, $j \leq i$. Also, we assume that the error terms $\{Z_i\}$ have a common continuous distribution function F_Z and stationary density function f_Z. As the error terms follow the model in (3.1), $\sigma(Z_{i-1}) > 0$ represents the unknown conditional variance function of Z_{i-1}. In practice, X_i may contain lagged values of Y_i and lagged values of other covariates, for example $X_i = Y_{i-1}$.

As an alternative to the least squares regression that employs information from the conditional mean of the dependent variable given a set of regression variables, X_i, quantile regression provides an appealing method, to estimate the model in (3.16), since it can give a more informative picture of the distribution by specifying different conditional quantile positions $\tau \in (0, 1)$ of the dependent variable given a set of predictors in the regression model [Koenker (2005)]. Thus, we use the quantile regression approach to estimate the parameters β in the model in (3.16).

In the context of linear regression models with heteroscedastic errors, parametric specification of the conditional variance of errors, $\sigma(\cdot)$, has been extensively studied in the literature. The use of two-step methods is often suggested for the estimation of the model. For instance, in the quantile regression for the linear heteroscedastic model investigated by Koenker and Zhao (1994) and Gutenbrunner and Jurečková (1992), $\sigma(X_i)$ is assumed to be a linear function of the regression variables, $\sigma(X_i) = X_i^T\gamma$, where γ is a k-dimensional vector, which can be estimated by regressing the absolute residuals from initial estimation on X_i. Koenker and Zhao (1996) discuss the case of quantile regression for linear parametric ARCH models, with $\sigma(\epsilon_i) = \epsilon_i^T\gamma$ as a linear function of the absolute residuals $\epsilon_i = (1, |Z_{i-1}|, \ldots, |Z_{i-q}|)$, where γ is a $(q+1)$-dimensional vector, which is computed based on the estimated residuals of the model. Likewise, for an example of the nonparametric specification of the conditional variance of errors, Zhao (2001) studies the median regression models in the presence of heteroscedasticity of unknown form, in the scenario of independent observations, where $\sigma(X_i)$ is an unknown function of the regression variables, which is estimated in a nonparametric way based on the estimated residuals of the model.

Let $F_Z^{-1}(\tau)$ be the τth quantile of the error term Z_i. The conditional quantile functions of Y_i in the model (3.16) given X_i, are of the form of

$$Q_{Y_i}(\tau \mid X_i) = X_i^T\beta + F_Z^{-1}(\tau) = X_i^T\beta(\tau), \tag{3.17}$$

where $\beta(\tau) = (\beta_0 + F_Z^{-1}(\tau), \beta_1, \ldots, \beta_{d-1})^T$. In many applications, conditional heteroscedasticity is often modeled based on the residuals with the effect of the location component removed [see, e.g., Engle (1982); Gutenbrunner and Jurečková (1992); Koenker and Zhao (1994, 1996) among many others]. For instance, in Engle (1982), the estimation of the ARCH regression model is performed in two steps: the first step involves estimating the autoregressive parameters by least squares and computing the residuals, and the second step consists in estimating the parameters of the ARCH component by regressing the squared residuals on the lagged squared residuals. Similarly, Koenker and Zhao (1996) use a two-step procedure to estimate the quantile regression for linear parametric ARCH models. In this section, we use a two-step procedure to estimate the model in (3.16) and analyze the behavior in the estimation of the limiting variance of $\hat{\beta}(\tau)$ when the proposed estimator $\hat{f}_Z$ is used.

The two-step strategy to estimate model (3.16) is as follows:

Step 1: Perform the conditional quantile estimation to obtain an estimator of the d-dimensional coefficient vector $\beta(\tau)$. Then, as Koenker and Bassett (1978), for any fixed τ, the τ-th regression quantile estimator of $\beta(\tau)$ is given by

$$\hat{\beta}(\tau) = \underset{b \in \mathbb{R}^d}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau \left(Y_i - X_i^T b \right), \tag{3.18}$$

where $\rho_\tau(u) = u\,(\tau - I(u < 0))$ is τ-th quantile loss, defined for $0 < \tau < 1$, which simply gives different weights to positive and negative values and $I(\cdot)$ denotes the indicator function. For example, for $\tau = 0.5$, the objective function (3.18) reduces to the recognized *Least Absolute Deviation* [Pollard (1991)] maximization problem: $\operatorname{argmin}_{b \in \mathbb{R}^d} \sum_{i=1}^{n} \left| Y_i - X_i^T b \right|$. The calculation of the regression quantiles by using standard linear programming techniques is quite efficient. For a thorough description of algorithmic aspects, see Koenker and D'Orey (1987). We use the function `qr` in the `R` package `quantreg` to get the estimates in (3.18) for different values of τ.

Step 2: Compute the estimated noises:

$$\hat{Z}_i = Y_i - X_i^T \hat{\beta}(\tau), \quad i = 1, \ldots, n., \tag{3.19}$$

and utilize the resulting $\hat{Z}_i$ values to estimate the conditional variance of the error terms, $\sigma(\cdot)$, by applying the widely used nonparametric kernel smoothing estimate $\hat{\sigma}^2(\cdot)$ given in (3.6). The quantity $\hat{Z}_i$ is specific to the τ-th quantile. To select the bandwidth h_n of the estimator $\hat{\sigma}^2(\cdot)$, as discussed in Subsection 3.2.1, use $h_n = 2h_n^*$, where h_n^* is the automatic plug-in bandwidth selector of Ruppert et al. (1995), which is implemented by using the command `dpill` in the `R` package `KernSmooth`.

It is well known [Koenker (2005)] that the limiting variance of $\hat{\beta}(\tau)$ in (3.18) depends on the reciprocal of the density function of the model error terms, evaluated at the quantile of interest, $1/f_Z(F_Z^{-1}(\tau))$, called the sparsity. Thus, to compute the estimation of the precision of the τ-th quantile regression estimate directly, the unknown quantity $f_Z(F_Z^{-1}(\tau))$ needs to be estimated, which is hard to do adequately. Then, an efficient estimate of the density, $\hat{f}_Z(\cdot)$, will give us a more accurate estimate of this limiting

variance to have the possibility to make inference about the estimation of $\hat{\beta}(\tau)$, for example, to construct confidence intervals. Now, based on Koenker et al. (2018), we give some regularity assumptions for establishing the asymptotic behavior of the quantile regression estimator, $\hat{\beta}(\tau)$.

Assumption 7. The density of the error terms f_Z is bounded and continuous, that $0 < f_Z\left(F_Z^{-1}(\tau)\right) < \infty$ for any $0 < \tau < 1$.

Assumption 8. (i) For each i, $1 \leq i \leq n$, Z_i is independent of the series values X_i, Z_{i-1}, X_{i-1}, Z_{i-2},.... (ii) Let $\{X_i\}$ be strictly stationary and ergodic process, with $E\left(\|X_t\|^2\right) < \infty$. (iii) Let $\{Z_i\}$ as in assumption 6(ii). (iv) There exist positive definite matrix Ω_0 such that $\lim_{n\to\infty} n^{-1} \sum X_i X_i^\top = \Omega_0$ and $\max_{i=1,\ldots,n} \|X_i\| / \sqrt{n} \to 0$.

Assumption 7 is a common assumption in quantile regression [see Koenker (2005)]. The dependence structure in Assumption 8 offers a very flexible framework to study asymptotic theory. Also, we focus on the model with stationary covariates; hence, Assumption 8(ii) assumes that $\{X_t\}$ is strictly stationary. Assumption 8(iv) is familiar throughout the literature on M-estimators for regression models [see Koenker (2005)]. Then, following the result according to Koenker et al. (2018), under Assumptions 7 and 8, for any fixed $\tau \in (0, 1)$, we have that the asymptotic behavior of $\hat{\beta}(\tau)$ in (3.18), in step 1, is given by

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) \Rightarrow N\left(0, \Gamma(\hat{\beta}(\tau))\right), \tag{3.20}$$

where the asymptotic variance $\Gamma(\hat{\beta}(\tau))$ is of the form of

$$\Gamma(\hat{\beta}(\tau)) = \frac{\tau(1-\tau)}{f_Z^2(F_Z^{-1}(\tau))}\Omega_0^{-1}. \tag{3.21}$$

From (3.21), as we mentioned above, the limiting variance $\Gamma(\hat{\beta}(\tau))$ depends on the quantity $f_Z(F_Z^{-1}(\tau))$. However, in practice, both the stationary density, f_Z, and the error terms Z_i are unknown, so we need to estimate these two quantities to have an estimation of $f_Z(F_Z^{-1}(\tau))$. Thus, for any fixed $\tau \in (0, 1)$, the procedure to estimate the sparsity is as follows:

(1) Compute $\hat{Z}_i$, for $i = 1, \ldots, n.$, and $\hat{\sigma}(\cdot)$ following steps 1 and 2 described above on the previous page in the two-step strategy.

(2) Estimate $F_Z^{-1}(\tau)$ by the sample τ-th quantile of $\hat{Z}_1, \ldots, \hat{Z}_n$, denoted by $\hat{F}_{\hat{Z}}^{-1}(\tau)$.

(3) Use $\hat{Z}_i$, $\hat{\sigma}(\cdot)$, and $\hat{F}_{\hat{Z}}^{-1}(\tau)$ in the proposed estimator $\hat{f}_Z(z)$, in (3.5), to estimate $f_Z(z)$. Thus, the estimation of $f_Z(F_Z^{-1}(\tau))$ is done through the proposed estimator $\hat{f}_{\hat{Z}}(\hat{F}_{\hat{Z}}^{-1}(\tau))$.

Finally, plug $\hat{f}_{\hat{Z}}(\hat{F}_{\hat{Z}}^{-1}(\tau))$ into $\Gamma(\hat{\beta}(\tau))$ in (3.21) to obtain $\hat{\Gamma}(\hat{\beta}(\tau))$. The performance of the proposed methodology to estimate the limiting variance of $\hat{\beta}(\tau)$ is examined, in sections 3.3.3 and 3.3.4, through both simulation data and an empirical analysis of the monthly inflation rate for the United States.

Given the properties of our estimator $\hat{f}_Z$, we conjecture that, although we use the estimate residuals, $\hat{Z}$, of the model (3.16), the resulting estimator $\hat{f}_{\hat{Z}}(z)$ in (3.5) is $\sqrt{n}$-consistent. However, it is difficult to prove this result and establish asymptotic normality. Therefore, we have decided to omit the proof for now, intending to report it in a future work. As a result, we shall not consider the consistency properties or any other large sample distribution for the estimator $\hat{f}_{\hat{Z}}(z)$. Instead, in Subsection 3.3.3, we report extensive simulation experiments that mimic various cases of practical interest to observe the performance of the proposed estimator compared with the traditional nonparametric kernel density estimator $\tilde{f}_{\hat{Z}}(z)$ in (3.2).

3.3 Numerical studies

In this Section, we will first investigate the performance of our density estimate compared with the traditional nonparametric kernel density estimate through a simulation study (see Subsection 3.3.1) and then apply our method to the S&P 500 Index data (refer to Subsection 3.3.2). In the second part, we will observe our proposed methodology's performance in estimating the limiting variance of the estimated coefficients in quantile regression with nonparametric ARCH errors through a simulation study (Subsection 3.3.3). Then, we will provide an empirical analysis of the monthly inflation rate of the United States as an illustration of this methodology (outlined in Subsection 3.3.4).

3.3.1 A simulation study for performance of $\hat{f}_Z(z)$

In this Subsection, we conduct a simulation study to examine the performance of the proposed estimator $\hat{f}_Z(z)$ in (3.5). Specifically, we compared our density estimate in (3.5) with the widely used nonparametric kernel density estimator $\tilde{f}_Z(z)$ in (3.2). We examine the effect of the sample size n and the use of different heteroscedastic models on the performance of the proposed density estimates. Our estimator uses the Gaussian kernel, which will also be used in all subsequent sections. In addition, the true density $f_Z(z)$ of Z is computed by numerically evaluating (3.3).

As mentioned in Subsection 3.2.1, since we are dealing with dependent data, we require the specific bandwidth $b_n = \tilde{\nu} b_n^*$ for some $\tilde{\nu} > 1$ in the traditional nonparametric kernel density estimator $\tilde{f}_Z(z)$. Based on our simulations, we observed that using only the rule-of-thumb bandwidth choice b_n^* tends to undersmooth the data due to the ignorance of the structure of dependence. Thus, for the purpose of improving the performance of $\tilde{f}_Z(z)$ in terms of mean-squared-error, we experimented with different bandwidths $b_n = \tilde{\nu} b_n^*$ for $\tilde{\nu} = 1.5, 1.8, 2$. We found that using $\tilde{\nu} = 2$ generates a better performance for this estimator under this context of models. Therefore, we use $b_n = 2b_n^*$ as the bandwidth for computing the density estimate $\tilde{f}_Z(z)$ in the simulation studies and applications from now. The rule-of-thumb bandwidth [Silverman (1986)] choice b_n^* is calculated using the `R` function `bw.nrd0`.

Consider observations Z_i, $1 \leq i \leq n$, from the following three models that impose different nonlinear structures on $\sigma(Z_{i-1})$, such as square-root heteroscedasticity (Model 1), absolute value heteroscedasticity (Model 2), and exponential-type heteroscedasticity (Model 3):

$$\text{Model 1:} \quad Z_i = \sigma(Z_{i-1})\, e_i = \sqrt{\alpha_0 + \alpha_1 Z_{i-1}^2}\, e_i, \quad \text{with } (\alpha_0, \alpha_1) = (0.1, 0.5); \tag{3.22}$$

$$\text{Model 2:} \quad Z_i = \sigma(Z_{i-1})\, e_i = \Big(\alpha_0 + \alpha_1 \mid Z_{i-1} \mid\Big) e_i, \quad \text{with } (\alpha_0, \alpha_1) = (0.1, 0.5); \tag{3.23}$$

$$\text{Model 3:} \quad Z_i = \sigma(Z_{i-1})\, e_i = \exp\Big(\alpha_0 + \alpha_1 Z_{i-1}\Big) e_i, \quad \text{with } (\alpha_0, \alpha_1) = (-1, 0.5) \tag{3.24}$$

For noise e_i, we assume that $e_i \overset{\text{iid}}{\sim} N(0,1)$. The normality assumption is quite common assumption in ARCH models. Here we have that Model 1 in (3.22) is the usual ARCH(1) model [Engle (1982)]. Model 2 in (3.23), the absolute value form, requires both parameters to be positive to assure a positive value for $\sigma(Z_{i-1})$ for all the observations i. Finally, Model 3 in (3.24), the exponential form, has the advantage that $\sigma(Z_{i-1})$ is positive for

all values of α_0 and α_1; however, the data generated from such a model have infinite variance for any value of $\alpha_1 \neq 0$, see Engle (1982). Hence, to avoid the issue of infinite variance in this model 3, we use a negative number for α_0 in the simulation.

For the simulation, we use five sample sizes $n = 50, 100, 200, 400, 1000$, ranging from small to medium sample sizes. Also, we define a "warmup" period equal to 500 to reach an approximate stationary start for the simulated time series. For $\hat{f}_Z(z)$ in (3.5), we compute the mean-squared-error (MSE) as

$$\text{MSE}(\hat{f}_Z) = \text{average of 1000 realizations of } \left\{ \frac{q_Z(0.975) - q_Z(0.025)}{100} \sum_{j=1}^{100} [\hat{f}_Z(z_j) - f_Z(z_j)]^2 \right\}.$$

In the last expression we have that $q_Z(0.025)$ and $q_Z(0.975)$ are the 2.5% and 97.5% quantile of Z, respectively, and $z_1, \ldots, z_{100}$ are equally spaced points on the interval $[q_Z(0.025), q_Z(0.975)]$. The MSE for $\tilde{f}_Z(z)$ is defined similarly. In order to make an easy comparison, we also compute the relative MSE (RMSE) of $\hat{f}_Z$ as RMSE = $\text{MSE}(\hat{f}_Z)/\text{MSE}(\tilde{f}_Z)$, where a value of RMSE ≤ 1 indicates a better performance of our proposed estimator $\hat{f}_Z(z)$.

Table 3.1. MSE of $\hat{f}_Z$ and relative MSE (RMSE) of the proposed estimator of f_Z relative to the nonparametric kernel density estimator $\tilde{f}_Z$, at different sample sizes n. Numbers of RMSE ≤ 1 indicate better performance of $\hat{f}_Z(z)$.

	n	MSE($\hat{f}_Z$)	RMSE
Model 1	50	0.0101	0.136
	100	0.0052	0.104
	200	0.0028	0.083
	400	0.0014	0.068
	1000	0.0006	0.061
Model 2	50	0.0318	0.152
	100	0.0165	0.117
	200	0.0092	0.098
	400	0.0051	0.092
	1000	0.0025	0.090
Model 3	50	0.0062	0.108
	100	0.0033	0.088
	200	0.0019	0.073
	400	0.0011	0.071
	1000	0.0006	0.080

Table 3.1 displays the MSE of $\hat{f}_Z$ in (3.5) and its RMSE to the nonparametric kernel density estimator $\tilde{f}_Z$ in (3.2). The following observations can be made from the

table: Firstly, as the sample size increases, the MSE of $\hat{f}_Z$ decreases; this indicates an increasingly better performance of the proposed density estimates. Secondly, using $\hat{f}_Z(z)$ substantially reduces MSE with RMSE around 90% (i.e., RMSE $\leq 10\%$) for most cases. For instance, the MSE reduction for a sample size of $n = 100$ is approximately 89% for all the models considered. Thirdly, the larger the sample size, the smaller the RMSE, suggesting a good relative advantage of using $\hat{f}_Z(z)$ over the estimator $\tilde{f}_Z(z)$. This is because both estimators have bias terms that decrease as the sample size increases. While theoretically negligible, these bias terms may have some effects in finite samples. Since $\text{MSE} = \text{Bias}^2 + \text{Variance}$, as bias decreases, the variance plays a dominating role, leading to increasingly better relative performance of $\sqrt{n}$-consistent $\hat{f}_Z(z)$ than $\sqrt{nb_n}$-consistent $\tilde{f}_Z(z)$, where $b_n \propto n^{-1/5}$ decreases with sample size. Finally, the relative advantage of $\hat{f}_Z(z)$ over $\tilde{f}_Z(z)$ is consistent across the three models considered.

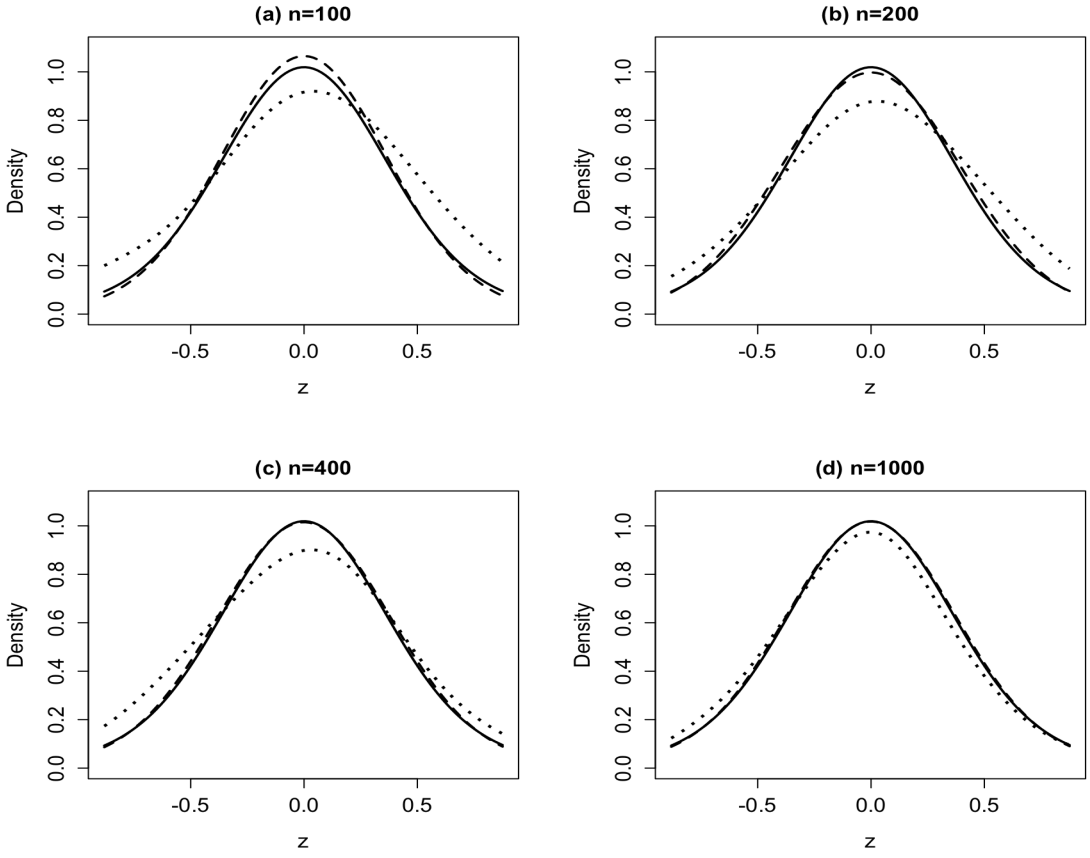

Figure 3.1. Four realizations (one realization on each figure) of density estimates with different sample sizes $n = 10, 200, 400, 1000$ from Model 1. In each plot, solid curve (——) is the true function, dotted curve ($\cdots$) is the usual nonparametric kernel estimates, and dashed curve ($--$) is the proposed method.

To illustrate the performance of $\hat{f}_Z$, Figure 3.1 plots four realizations of density estimations from Model 1 with different sample sizes. In all the sample sizes considered, $\hat{f}_Z$ (dashed curve in all plots) matches very well the true densities (solid curve in all plots) for the realizations. On the other hand, there is an inferior performance for $\tilde{f}_Z$ (dotted curve in all plots) for a sample size of 100 and 200. When the sample size increases, $\tilde{f}_Z$ works reasonably well for density estimate. This behavior of poor performance of the usual nonparametric kernel density estimator, $\tilde{f}_Z$, occurs relatively often in our simulations.

3.3.2 Application of $\hat{f}_Z(z)$ to the S&P 500 index data

For this application, we use the proposed procedure to estimate the marginal density of the weekly log returns for Standard & Poor 500, hereafter S&P 500, stock market weekly index (available from `https://finance.yahoo.com/quote/%5ESPX/history?p=%5ESPX`), which has been of particular interest in financial econometrics. See Chan et al. (2013) or Zhao and Wu (2008), for example. The S&P 500 index monitors the stock performance of 500 of the largest companies listed on the United States stock exchanges. Thus, it is a widely followed equity index that allows investors to evaluate market performance by comparing the current stock price levels with past prices. Since we use weekly frequency data, we do not encounter the issue of missing data points, as is the case of the daily S&P 500 index. This fact over the daily index is due to stocks not being traded on weekends or holidays, only on so-called trading days. This results in missing data points in the daily time series, which typically require estimation before analysis.

The data set consists of 451 weekly observations, $S_0, S_1, \ldots, S_{450}$, of S&P 500 index from 5 January 2015 to 21 August 2023. We work with the centered weekly log returns (as percentages) for S&P 500 index, that is, $Z_i = r_i - n^{-1}\sum_{i=1}^{n} r_i$, with $r_i = 100(\log(S_i) - \log(S_{i-1}))$ for $i = 1, \ldots, 450$. Ding et al. (1993) analyzed the daily log return series Z_i of the S&P 500 index, spanning from 3 January 1928 to 30 August 1991. Their findings revealed that the S&P 500 returns Z_i are not independent and identically distributed. Moreover, they observed that Z_i exhibit some short-term memory and that there exists a small amount of predictability in stock returns. For additional insights on the serial dependence among S&P 500 returns, see, for example, Caporale and Gil-Alana (2004) and Awartani and Corradi (2005) and the references therein.

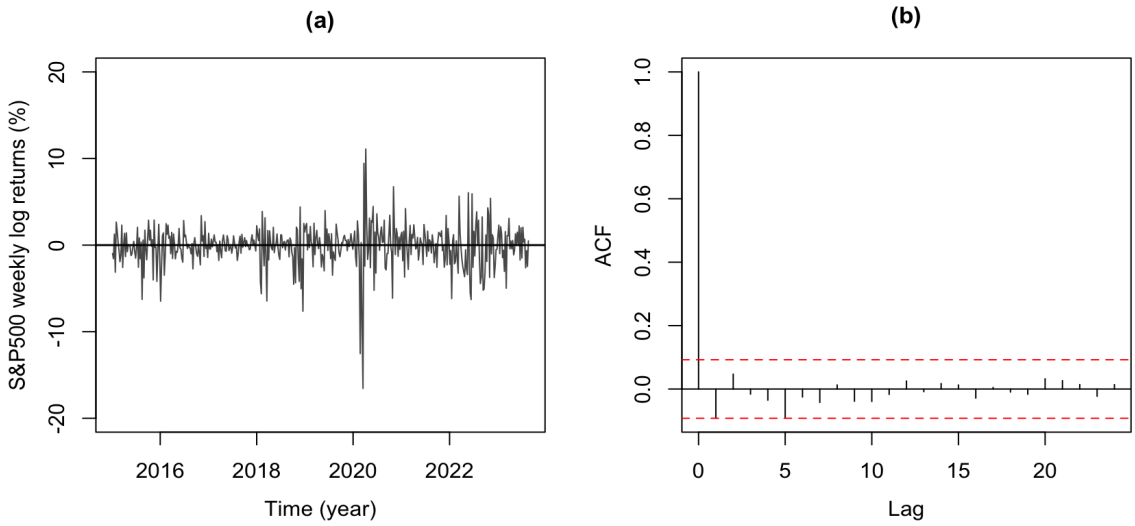

Figure 3.2. (a) The S&P 500 weekly log returns during 5 January 2015 - 21 August 2023, and (b) The estimated sample autocorrelation function (ACF) for the S&P 500 weekly log returns.

In order to analyze the behavior of the S&P 500 weekly log returns Z_i, we use Figure 3.2. From Figure 3.2(a), we can see that: (i) the time series is rather stable around zero. In fact, as the stationarity test of Kwiatkowski et al. (1992) (`R` command `kpss.test` in package `tseries`) has a p-value of 0.1, the null hypothesis of stationarity is not rejected at a level of significance 0.05; and (ii) the volatility of log returns is observed to change over time, where in particular, there are one period, specifically around 2020, where these volatilities are much higher than any other time frame considered here. Additionally, there is evidence of volatility clustering. Figure 3.2(b) plots the estimated sample autocorrelation function (ACF) for the S&P 500 weekly log returns, where the dashed red lines show the 95% confidence interval for these estimated sample ACF. We observe that this time series has no serial correlation, indicating the absence of a model of the conditional mean for the process. However, upon analyzing the ACF of the squared log returns (which is not shown here), we notice some significant autocorrelations. Thus, there is a higher-order serial dependence structure in the log returns, verifying the existence of ARCH effects in the time series.

The analysis from Figure 3.2 previously given suggests that a suitable model for this dataset should possess a volatility structure that varies with time, as indicated by the nonparametric model in (3.1). Hence, the proposed estimator $\hat{f}_Z(z)$ in (3.5) can be applied to estimate the marginal density for the S&P 500 weekly log returns.

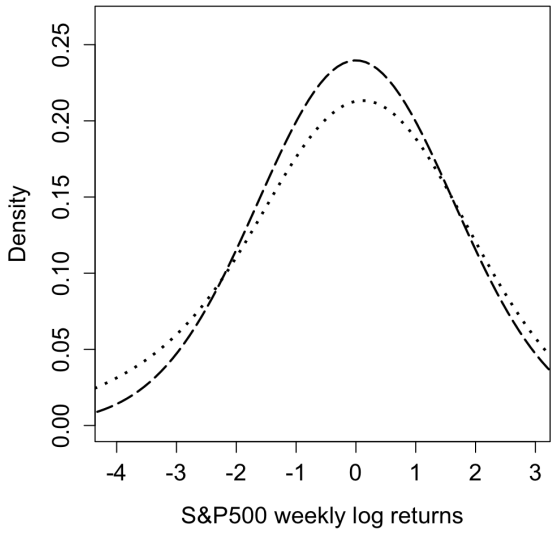

Figure 3.3. The estimated density curves of the S&P 500 weekly log returns. The dotted curve $(\cdots)$ is the usual nonparametric kernel estimate, and the dashed curve $(--)$ is the proposed method.

Figure 3.3 presents the estimated density curves for the S&P 500 weekly log returns using both the usual nonparametric kernel density estimate $\tilde{f}_Z(z)$ in (3.2) (dotted curve), and our density estimate $\hat{f}_Z(z)$ in (3.5) (dashed curve) for comparison. The density curves are evaluated at 100 equally spaced points between the 2.5% and 97.5% quantiles of the log returns in the data. The two estimated density curves seem to have a similar shape for log returns. However, $\tilde{f}_Z(z)$ overestimates the low and high log returns and underestimates the range of middle log returns compared to our density estimate $\hat{f}_Z(z)$. This behavior is similar to what we observed in Figure 3.1 for the specific realizations of density estimates in the simulation study in Subsection 3.3.1.

3.3.3 A simulation study for performance of $\hat{f}_Z(z)$ in the context of Quantile regression

In this Subsection, we conduct an extensive simulation study designed to investigate the performance of our proposed density estimate $\hat{f}_Z$ in the context of quantile regression with heteroskedastic form for the error term. Specifically, we compared our density estimate in (3.5) with the widely used nonparametric kernel density estimator $\tilde{f}_Z(z)$ in (3.2) to estimate the limiting variance of the estimated coefficients in a quantile model with different heteroskedastic forms for the error term noticed by Z.

Each model under consideration is of the form

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + Z_i, \quad i = 1, \ldots, n. \tag{3.25}$$

For the error term Z_i, similar to Subsection 3.3.1, we consider three widely used forms of heteroscedasticity to set the three models to study:

$$\text{Form 1:} \quad Z_i = \sigma(Z_{i-1})\, e_i = \sqrt{\alpha_0 + \alpha_1 Z_{i-1}^2}\, e_i, \quad \text{with } (\alpha_0, \alpha_1) = (0.1, 0.5); \tag{3.26}$$

$$\text{Form 2:} \quad Z_i = \sigma(Z_{i-1})\, e_i = \Big(\alpha_0 + \alpha_1 \mid Z_{i-1} \mid\Big) e_i, \quad \text{with } (\alpha_0, \alpha_1) = (0.1, 0.5); \tag{3.27}$$

$$\text{Form 3:} \quad Z_i = \sigma(Z_{i-1})\, e_i = \exp\Big(\alpha_0 + \alpha_1 Z_{i-1}\Big) e_i, \quad \text{with } (\alpha_0, \alpha_1) = (-1, 0.5). \tag{3.28}$$

The noise e_i is taken to be $e_i \overset{\text{iid}}{\sim} N(0,1)$, the true parameter is set to be $\beta = (\beta_0, \beta_1, \beta_2)^T = (10, -4, 2)^T$, and the covariates X_1 and X_2, in vector $X_i = (1, X_{1i}, X_{2i})$, are generated by [see Zhao (2001)]:

$$X_1 = U + 0.2V, \quad X_2 = 0.2U + V, \quad U \sim N(5, 9), \quad V \sim U(0, 4).$$

Thus, we have the following three distinct models.

Model 1: This model is formed by the linear regression model in (3.25) and the square-root heteroscedasticity form 1 in (3.26).

Model 2: This model resembles Model 1, but with the absolute value heteroscedasticity form 2 in (3.27).

Model 3: This model also employs the linear regression model in (3.25) but involves the exponential-type heteroscedasticity form 3 in (3.28). As we mentioned in Subsection 3.3.1, this heteroscedasticity 3 is positive for all values of α_0 and α_1; however, the data generated from such a model have infinite variance for any value of $\alpha_1 \neq 0$, see Engle (1982). Hence, to avoid the issue of infinite variance in this heteroscedasticity 3, we use a negative number for α_0 in the simulation.

The estimation of each of the three models, for fixed value $\tau \in (0,1)$, is done by using the two-step strategy described in Subsection 3.2.2. This is, for each model, in step 1, we compute the estimation of $\beta(\tau) = (\beta_0 + F_Z^{-1}(\tau), \beta_1, \beta_2)^T$ using the quantile regression estimator in (3.18) (`R` command `qr` in package `quantreg`). Now, in step 2, we calculate

the regression residuals using $\hat{Z}_i = Y_i - X_i^T\hat{\beta}(\tau)$, for $i = 1, \ldots, n$. To ensure a fair comparison among the density estimators, it is necessary to center $\hat{Z}_i$ around its mean. This is because these estimated residuals do not have zero mean, but by construction, the stationary density of the true residuals has. Therefore, we use this approach for this simulation and the application in the following Subsection. Henceforth, the centered estimated residuals will be referred to as $\hat{Z}_i$.

Now, based on $\hat{Z}_i$: (i) we use the kernel smoothing estimate given in (3.6) to estimate the heteroskedastic form $\sigma(\cdot)$ and (ii) the proposed estimator $\hat{f}_{\hat{Z}}$ to estimate f_Z. The quantity $\hat{Z}_i$ is specific to the τ-th quantile. For each of the three models considered, we are interested in estimating the limiting variance $\hat{\Gamma}(\hat{\beta}_j(\tau))$, for $j = 0, 1, 2$, in (3.21). To do this, based on $\hat{Z}_1, \ldots, \hat{Z}_n$, we compare the following two methods:

(1) *OM*: (Our method) We utilize the proposed estimator $\hat{f}_{\hat{Z}}(z)$ evaluated at the quantile of interest $\hat{z}_\tau = \hat{F}^{-1}_{\hat{Z}}(\tau)$, to estimate the quantity $1/f_Z(F_Z^{-1}(\tau))$ present in the expression of $\Gamma(\hat{\beta}(\tau))$ in (3.21). Then, we plug $1/\hat{f}_{\hat{Z}}(\hat{z}_\tau)$ into $\Gamma(\hat{\beta}(\tau))$ to obtain $\hat{\Gamma}(\hat{\beta}(\tau))$.

(2) *NKDE*: Simalar as OM, but in this case, we use the nonparametric kernel density estimator $\tilde{f}_{\hat{Z}}(z)$, in (3.2), evaluated at the quantile of interest $\hat{z}_\tau$, to estimate $1/f_Z(F_Z^{-1}(\tau))$. Then, we plug $1/\tilde{f}_{\hat{Z}}(\hat{z}_\tau)$ into $\Gamma(\hat{\beta}(\tau))$ to obtain $\hat{\Gamma}_{\text{NKDE}}(\hat{\beta}(\tau))$.

We performed a simulation study to compare the two methods mentioned above in order to estimate the limiting variance of $\hat{\beta}_j(\tau)$ under the three models considered; the number of replications in each method is $m = 1000$. For each run, we use sample sizes of $n = 100, 400$, and 1000. We consider five quantile levels: 0.1, 0.25, 0.5, 0.75, and 0.9. The results for these sample sizes are the only ones we are presenting, as we acquired similar results for sample sizes of 50 and 200. The mean-squared-error (MSE) is a commonly used criterion to evaluate the performance of estimators. Thus, for any $0 < \tau < 1$ and using the proposed density estimator $\hat{f}_Z(z)$ in (3.5) based on the estimated residuals of the model $\hat{Z}_1, \ldots, \hat{Z}_n$ (OM method), we compute the MSE of the estimated asymptotic variance of $\hat{\beta}_j(\tau)$, for $j = 0, 1, 2$, as

$$\text{MSE}\left(\hat{\Gamma}(\hat{\beta}_j(\tau))\right) = \frac{\sum_{i=1}^{m}\left[\hat{\Gamma}^{(i)}(\hat{\beta}_j(\tau)) - \Gamma(\hat{\beta}_j(\tau))\right]^2}{m},$$

where $\hat{\Gamma}^{(i)}(\hat{\beta}_j(\tau))$ is the estimated asymptotic variance of $\hat{\beta}_j(\tau)$ on the i-th sample and

$\Gamma(\hat{\beta}_j(\tau))$ is the true asymptotic variance of $\hat{\beta}_j(\tau)$, respectively. The $\Gamma(\hat{\beta}_j(\tau))$ value is the diagonal value obtained from the matrix $\Gamma(\cdot)$ in (3.21). The MSE for the estimation of the asymptotic variance of $\hat{\beta}_j(\tau)$ by utilizing the nonparametric kernel density estimator (NKDE method) in (3.2) is defined similarly, this MSE is represented by $\text{MSE}\left(\hat{\Gamma}_{\text{NKDE}}(\hat{\beta}_j(\tau))\right)$. In order to make the results more readable, we also compute the relative MSE (RMSE), defined as

$$\text{RMSE}\left(\hat{\Gamma}(\hat{\beta}_j(\tau))\right) = \frac{\text{MSE}\left(\hat{\Gamma}(\hat{\beta}_j(\tau))\right)}{\text{MSE}\left(\hat{\Gamma}_{\text{NKDE}}(\hat{\beta}_j(\tau))\right)},$$

with RMSE ≤ 1 indicating a better performance in estimating the asymptotic variance of $\hat{\beta}_j(\tau)$, for $j = 0, 1, 2$, using our proposed method OM. The simulation results are summarized in Table 3.2.

Table 3.2 shows the MSE ($\times 100$) of the estimated asymptotic variance $\hat{\Gamma}(\hat{\beta}_j(\tau))$ by utilizing the proposed method (OM) and the RMSEs. We make the following observations:

(1^0) In quantiles close to $\tau = 0.5$ (specifically for τ=0.25, 0.50, and 0.75), the use of our method (OM) results in a substantial reduction in MSE with RMSE below 91% for most cases considered. For example, for $\hat{\beta}_1$ in Model 1 with $\tau = 0.5$ and $n = 100$, the reduction in the MSE is about 84%. For values of τ of 0.1 and 0.9, when we have a small sample size $n = 100$, the NKDE method performs better than our method (OM). However, as the sample size increases to 400, the OM method shows a reduction in MSE with an RMSE of around 36% for most cases. This reduction in MSE for our method is even better for $n = 1000$.

(2^0) The performance of our method (OM) to estimate $\hat{\Gamma}(\hat{\beta}_j(\tau))$, for $j = 0, 1, 2$, is consistent across the three models and sample sizes considered, and it is even better in the cases of values of τ close to 0.5. This may be because more observations are available as τ gets closer to the center.

(3^0) The larger the sample size, the smaller the RMSE, and thus there is a relative advantage of using $\hat{f}_{\hat{Z}}(z)$ over $\tilde{f}_{\hat{Z}}(z)$ to estimate the asymptotic variance of $\hat{\beta}_j(\tau)$'s.

In conclusion, substantial efficiency gain can be achieved by using our proposed density estimator $\hat{f}_{\hat{Z}}(z)$ in (3.5) (OM method), based on the estimated residuals of the model, to estimate the asymptotic variance of $\hat{\beta}_j(\tau)$'s in the quantile regression model compared with the NKDE method.

Table 3.2. MSE (×100) of the estimation for the asymptotic variance of $\hat{\beta}_j(\tau)$, for $j = 0, 1, 2$, by utilizing the proposed estimator $\hat{f}_{\hat{Z}}(z)$ (OM method) and relative MSE (RMSE). RMSE ≤ 1 indicates a better performance in estimating the asymptotic variance of $\hat{\beta}_j(\tau)$, using $\hat{f}_{\hat{Z}}(z)$.

			MSE (×100)			RMSE		
	n	τ	$\hat{\Gamma}(\hat{\beta}_0(\tau))$	$\hat{\Gamma}(\hat{\beta}_1(\tau))$	$\hat{\Gamma}(\hat{\beta}_2(\tau))$	$\hat{\Gamma}(\hat{\beta}_0(\tau))$	$\hat{\Gamma}(\hat{\beta}_1(\tau))$	$\hat{\Gamma}(\hat{\beta}_2(\tau))$
Model 1	100	0.10	303.0325	0.1626	4.3804	1.171	1.245	1.185
		0.25	35.5103	0.0209	0.4688	0.361	0.372	0.353
		0.50	22.6300	0.0126	0.2791	0.152	0.160	0.141
		0.75	45.2515	0.0241	0.5827	0.391	0.396	0.375
		0.90	298.5324	0.1593	4.2161	1.123	1.146	1.103
	400	0.10	93.5466	0.0509	1.3160	0.771	0.789	0.742
		0.25	9.1675	0.0048	0.1237	0.345	0.353	0.340
		0.50	4.1599	0.0023	0.0546	0.113	0.117	0.104
		0.75	8.8603	0.0046	0.1140	0.368	0.362	0.348
		0.90	81.6479	0.0410	1.1311	0.633	0.633	0.606
	1000	0.10	34.9604	0.0181	0.4972	0.506	0.505	0.490
		0.25	3.3439	0.0018	0.0463	0.332	0.342	0.334
		0.50	1.8298	0.0010	0.0222	0.098	0.104	0.086
		0.75	3.5278	0.0019	0.0459	0.356	0.363	0.343
		0.90	39.3961	0.0205	0.5650	0.494	0.492	0.474
Model 2	100	0.10	9.8656	0.0053	0.1352	1.162	1.197	1.128
		0.25	1.1899	0.0007	0.0152	0.386	0.388	0.374
		0.50	0.6416	0.0004	0.0080	0.172	0.179	0.159
		0.75	1.0435	0.0006	0.0140	0.393	0.392	0.376
		0.90	8.9040	0.0051	0.1305	1.085	1.159	1.070
	400	0.10	2.8060	0.0014	0.0396	0.679	0.667	0.656
		0.25	0.2510	0.0001	0.0033	0.392	0.386	0.384
		0.50	0.1362	0.0001	0.0017	0.132	0.134	0.118
		0.75	0.2481	0.0001	0.0031	0.413	0.430	0.406
		0.90	2.8315	0.0015	0.0405	0.632	0.628	0.611
	1000	0.10	1.3435	0.0007	0.0194	0.552	0.552	0.540
		0.25	0.0919	0.0001	0.0012	0.376	0.396	0.381
		0.50	0.0550	0.0001	0.0007	0.116	0.119	0.110
		0.75	0.0969	0.0001	0.0013	0.412	0.395	0.394
		0.90	1.4433	0.0007	0.0210	0.549	0.549	0.540
Model 3	100	0.10	124.0774	0.0638	1.6628	1.153	1.189	1.124
		0.25	17.5952	0.0097	0.1992	0.343	0.352	0.303
		0.50	10.3859	0.0061	0.1347	0.132	0.146	0.125
		0.75	21.4572	0.0107	0.2261	0.378	0.364	0.324
		0.90	136.5286	0.0738	1.8223	1.157	1.238	1.147
	400	0.10	30.0570	0.0154	0.4176	0.618	0.621	0.595
		0.25	4.0737	0.0023	0.0486	0.288	0.315	0.270
		0.50	3.0049	0.0014	0.0336	0.122	0.119	0.102
		0.75	3.9036	0.0021	0.0471	0.291	0.312	0.272
		0.90	33.2228	0.0182	0.4591	0.582	0.589	0.548
	1000	0.10	13.9714	0.0072	0.1966	0.529	0.525	0.507
		0.25	1.6179	0.0009	0.0195	0.257	0.261	0.228
		0.50	1.3688	0.0008	0.0166	0.120	0.127	0.104
		0.75	1.4945	0.0008	0.0183	0.248	0.267	0.228
		0.90	17.2736	0.0088	0.2454	0.505	0.506	0.490

In Subsection 3.2.2, it was mentioned that the estimation of $\Gamma(\hat{\beta}(\tau))$ in (3.21) depends on the reciprocal of the density function of the model error terms Z, evaluated at the quantile of interest $z_\tau = F_Z^{-1}(\tau)$, which would be $1/f_Z^2(z_\tau)$. This means that the accuracy of this estimation $\hat{\Gamma}(\hat{\beta}(\tau))$ varies based on the accuracy of the point estimation $f_Z(z_\tau)$. So, even though, in some cases, the estimation of the limiting variance appears to be good, a bias in the estimation of $f_Z(z_\tau)$ can lead to incorrect conclusions about the quality of the estimation of $\Gamma(\hat{\beta}(\tau))$. To investigate the behavior of the point estimation of $f_Z(z_\tau)$, we do the following: First, we compute a ratio using the formula Ratio$= \hat{f}_{\hat{Z}}(\hat{z}_\tau)/\tilde{f}_{\hat{Z}}(\hat{z}_\tau)$, with $\hat{z}_\tau = \hat{F}_{\hat{Z}}^{-1}(\tau)$. This is done in Table 3.3. Second, based on the expression (3.15), we calculate an approximation of bias for the nonparametric kernel density estimator $\tilde{f}_{\hat{Z}}$ in Table 3.4.

Table 3.3 displays the ratio, described in the last paragraph, at different sample sizes and levels of τ for the three models considered. A value of Ratio>1 indicates that $\tilde{f}_{\hat{Z}}(\hat{z}_\tau) < \hat{f}_{\hat{Z}}(\hat{z}_\tau)$. The reported values of the ratio are based on averaging over the $m = 1000$ realizations of the simulation study described above. It can be seen that for a small sample size ($n = 100$) and quantiles close to $\tau = 0.5$, as Ratio>1, the point estimation using the nonparametric kernel density $\tilde{f}_{\hat{Z}}(\hat{z}_\tau)$ tends to be less than estimation $\hat{f}_{\hat{Z}}(\hat{z}_\tau)$. Also, when $n = 100$ and we consider quantiles $\tau = 0.1$ and $\tau = 0.9$, we have Ratio<1, indicating that $\tilde{f}_{\hat{Z}}(\hat{z}_\tau) > \hat{f}_{\hat{Z}}(\hat{z}_\tau)$. This behavior in the ratio values might be due to the bias present in the estimator $\tilde{f}_{\hat{Z}}$, which diminishes when the sample size increases, as we see for the sample size of 1000 where the ratio is approximately equal to 1 in all cases. Thus, in Table 3.4, we compute an approximation of this bias of $\tilde{f}_{\hat{Z}}$.

From the expression of bias for the nonparametric kernel density estimator $\tilde{f}_Z(z)$ in Corollary 2, we have that $\mathrm{C}_Z(z)$ in (3.15) represents a constant coefficient term, with respect to n, for the bias of the estimator $\tilde{f}_Z(z)$. The term $\mathrm{C}_Z(z)$ ignores the calculation of bandwidth b_n, which determines the bias for specific sample sizes. We do this to have a clear idea of the direction of the bias of $\tilde{f}_Z(z)$ and to observe for what values of τ this bias shows a bigger magnitude. Thus, for the three models considered above, we get a theoretical approximation for the bias of $\tilde{f}_Z$ through the approximation of $\mathrm{C}_Z(z)$ in (3.15). From now, we refer to $\mathrm{C}_Z(z)$ as bias approximation of $\tilde{f}_Z(z)$.

Table 3.4 presents the bias approximation $\mathrm{C}_Z(z)$ evaluated at specific quantiles for the three models considered, which possess different forms of heteroscedasticity. To compute this approximation, we numerically evaluate (3.15). Each value z_τ in Table 3.4

Table 3.3. Ratio= $\hat{f}_{\hat{Z}}(\hat{z}_\tau)/\tilde{f}_{\hat{Z}}(\hat{z}_\tau)$, with $\hat{z}_\tau = \hat{F}_{\hat{Z}}^{-1}(\tau)$, at different sample sizes and different levels of τ for the three models considered. Ratio>1 indicates that $\tilde{f}_{\hat{Z}}(\hat{z}_\tau) < \hat{f}_{\hat{Z}}(\hat{z}_\tau)$.

		Ratio		
n	τ	Model 1	Model 2	Model 3
100	0.10	0.896	0.903	0.911
	0.25	1.124	1.124	1.109
	0.50	1.218	1.203	1.188
	0.75	1.125	1.121	1.103
	0.90	0.894	0.902	0.904
400	0.10	0.921	0.927	0.940
	0.25	1.082	1.076	1.069
	0.50	1.146	1.135	1.118
	0.75	1.082	1.080	1.067
	0.90	0.919	0.926	0.940
1000	0.10	0.940	0.948	0.961
	0.25	1.058	1.056	1.048
	0.50	1.105	1.094	1.080
	0.75	1.059	1.059	1.048
	0.90	0.939	0.950	0.961

corresponds to the τ-th quantile of error terms Z in each case. It is possible to observe that across all three models, the estimator $\tilde{f}_Z$ shows the following facts: Firstly, it tends to overestimate (positive bias) the true density f_Z for quantiles around $\tau \approx 0.1$ and $\tau \approx 0.9$. Here, it is useful to note that given the heteroscedasticity structure imposed on Z, large negative values of Z are associated with $\tau \approx 0.1$, and large positive values of Z are associated with $\tau \approx 0.9$. Secondly, the estimator $\tilde{f}_Z$ underestimates (negative bias) f_Z for quantiles close to 0.5 where $z_\tau = F_Z^{-1}(\tau) = 0$. Finally, as the quantile level τ approaches 0.5, the bias increases in magnitude.

The bias approximation pattern of $\tilde{f}_Z$ in Table 3.4 can explain the behavior of the ratio values in Table 3.3. This is, in Table 3.3, when τ is around values close to 0.5, we will tend to have a negative bias of $\tilde{f}_Z$, causing in some cases that $\tilde{f}_Z(\hat{z}_\tau) < \hat{f}_Z(\hat{z}_\tau)$. This results in a Ratio= $\hat{f}_{\hat{Z}}(\hat{z}_\tau)/\tilde{f}_{\hat{Z}}(\hat{z}_\tau) > 1$. On the other hand, when τ is around values close to 0.1 or 0.9, $\tilde{f}_Z$ will show a positive bias causing in some cases that $\tilde{f}_Z(\hat{z}_\tau) > \hat{f}_Z(\hat{z}_\tau)$. This generates a Ratio$<1$.

Based on Tables 3.3 and 3.4, we can observe two specific cases for small sample sizes ($n = 100$). Firstly, if τ is close to 0.5 (i.e. $z \approx 0$), then the bias of $\tilde{f}_Z(z)$ is negative and $\tilde{f}_Z(\hat{z}_\tau) < \hat{f}_Z(\hat{z}_\tau)$ for most of the replications in the simulation study. As a result, for $j = 0, 1, 2$, $\hat{\Gamma}_{\text{NKDE}}(\hat{\beta}_j(\tau)) > \hat{\Gamma}(\hat{\beta}_j(\tau))$ in most cases. Thus in this scenario, using $\tilde{f}_Z$ tends to overestimate the limiting variance $\Gamma(\hat{\beta}_j(\tau))$. Secondly, if $\tau \approx 0.1$ or $\tau \approx 0.9$,

Table 3.4. Approximation of bias of the widely used nonparametric kernel density estimator $\tilde{f}_Z(z)$ in (3.5), at different quantiles τ, $z_\tau = F_Z^{-1}(\tau)$, for the three models considered.

	τ	z_τ	$\mathrm{C}_Z(z_\tau)$
	0.10	-0.524	1.392
	0.25	-0.266	-1.228
Model 1	0.50	0.000	-3.787
	0.75	0.267	-1.204
	0.90	0.526	1.400
	0.10	-0.208	25.759
	0.25	-0.104	-16.924
Model 2	0.50	0.000	-70.500
	0.75	0.104	-16.484
	0.90	0.209	25.820
	0.10	-0.475	1.541
	0.25	-0.245	-1.522
Model 3	0.50	0.000	-4.728
	0.75	0.245	-1.513
	0.90	0.480	1.563

then the bias of $\tilde{f}_Z(z)$ is positive and $\tilde{f}_Z(\hat{z}_\tau) > \hat{f}_Z(\hat{z}_\tau)$ in most of the replications in the simulation study. In this case, we will observe that $\hat{\Gamma}_{\mathrm{NKDE}}(\hat{\beta}_j(\tau)) < \hat{\Gamma}(\hat{\beta}_j(\tau))$, which indicates a tendency to underestimate the limiting variance $\Gamma(\hat{\beta}_j(\tau))$ when using $\tilde{f}_Z$. These two specific cases can try to explain the pattern we observe in the performance of the confidence intervals for $\hat{\beta}(\tau)$ in Tables 3.5 and 3.6 below. An example of this behavior is illustrated in Figure 3.4.

To illustrate the performance of $\hat{f}_{\hat{Z}}$ for estimating the true density of model residuals f_Z, Figure 3.4 plots six realizations of density estimations of f_Z based on the estimated residuals from Models 1 and 2 with $n = 100$, and $\tau = 0.1, 0.5$ and 0.9. First, we observe that the proposed method (dashed curve in all plots) closely matches the true curves of the residuals (solid curve in all plots) somewhat well for the six realizations. By contrast, the nonparametric kernel density estimate $\tilde{f}_{\hat{Z}}$ (dotted curve in all plots) performs poorly in most realizations. However, this estimator $\tilde{f}_{\hat{Z}}$ works reasonably well when the sample size is increased (not shown in the plot) but not as well as our method. Second, we get a more accurate estimate for the quantity $f_Z(z_\tau)$ (the circle symbol in the solid curve in all plots) by using our proposed method (the square symbol in the dashed curve in all plots) compared with the estimation of this quantity using $\tilde{f}_{\hat{Z}}$ (the triangle symbol in the dotted curve in all plots) for the six realizations. It is noteworthy that the poor performance in the estimation of $f_Z(z_\tau)$ when we use $\tilde{f}_{\hat{Z}}$ can be due to the bias present in this estimator. This is, for small sample sizes, $\tilde{f}_{\hat{Z}}$ tends to overestimate $f_Z(z_\tau)$ at quantiles close to $\tau = 0.1$ or 0.9, and underestimate it at values of τ close to 0.5 as

described above in the last paragraph.

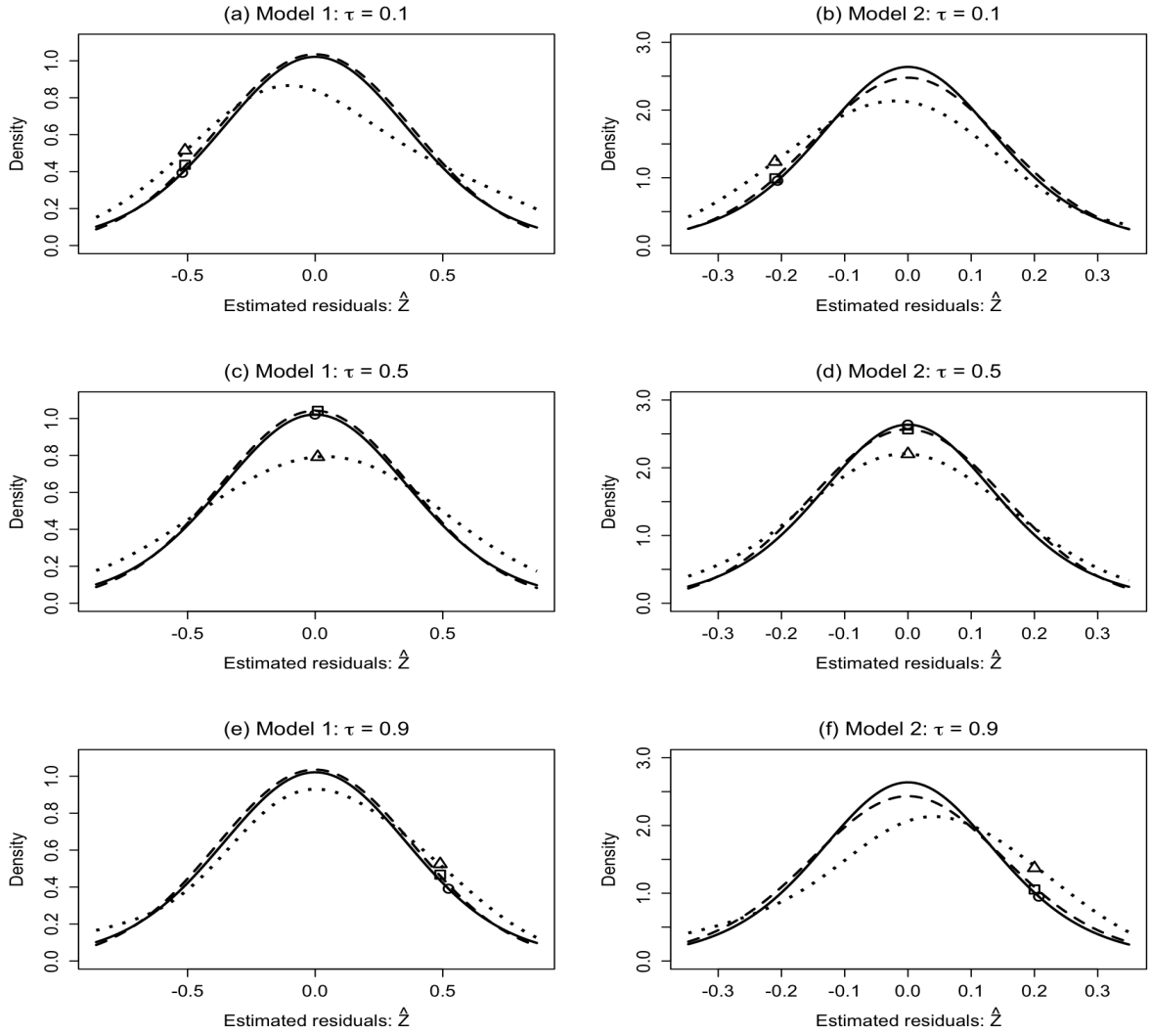

Figure 3.4. Six realizations (one realization on each figure) of density estimates for the estimated residuals from Models 1 and 2 with $\tau = 0.1, 0.5, 0.9$ and sample size $n = 100$. In each plot, solid curve (——) is the true function with the circle symbol ($\circ$) representing the true value of quantity $f_Z(z_\tau)$, where $z_\tau = F_Z^{-1}(\tau)$, dotted curve ($\cdots$) is the usual nonparametric kernel estimates with the triangle symbol ($\triangle$) representing the estimation of $f_Z(z_\tau)$ using this method (NKDE), and dashed curve ($--$) is the proposed method with the square symbol ($\square$) representing the estimation of $f_Z(z_\tau)$ using our method (OM)

To conclude this Subsection, for the simulation study, we present a confidence interval performance for $\hat{\beta}_0(\tau)$, $\hat{\beta}_1(\tau)$, and $\hat{\beta}_2(\tau)$ when we estimate the corresponding limiting variance using the two methods described earlier: (i) Our proposed method (OM) using the $\hat{f}_Z$, in (3.5), and (ii) the method NKDE using the estimator $\tilde{f}_Z$, in (3.2). Table 3.5

and Table 3.6 report results of coverage probability and lengths of confidence intervals, respectively, under the two methods to estimate the limiting variance $\hat{\Gamma}(\hat{\beta}_j(\tau))$, for $j = 0, 1, 2$. We exclusively focus on five quantile levels (τ) under the models mentioned above for $n = 100, 400$ and 1000. In all cases, we present results for confidence intervals that are intended to have a coverage of 0.95. The reported estimated coverage probability (see Table 3.5) and lengths (see Table 3.6) for the estimated confidence interval are based on averaging over the 1000 realizations of this simulation study.

The results presented in Table 3.5 indicate that the average 95% confidence interval coverage rate is a bit low for the smallest sample size ($n = 100$) in both methods but gets closer to the nominal 0.95 level as n increases for all quantiles, models, and estimated parameters. For quantiles $\tau = 0.25, 0.5$ and 0.75, both methods perform quite well as the empirical coverage probability is close to 95%. However, in most cases, our method (OM) performs little better than the NPKDE method. For quantiles $\tau = 0.1$ and 0.9, our approach (OM) outperforms the NPKDE method in all $\hat{\beta}(\tau)$'s. For example, under Model 1, for the confidence interval of $\hat{\beta}_1(\tau)$, with $n = 100$ and $\tau = 0.1$, our method (OM) exhibits an empirical coverage of 93%, whereas the NPKDE method only has a coverage rate of 90%. The relative advantage of the proposed method (OM) over the NPKDE approach is consistent across the three models, which consider different forms of heteroscedasticity.

From Table 3.6, we make the following observations:

(1^0) It can be seen that the mean lengths of confidence intervals are similar for both approaches when considering $\tau = 0.25, 0.5$, and 0.75. However, the proposed method (OM) generally exhibits lower standard errors for these average lengths than the NPKDE method. For instance, in Model 1 with $n = 100$ and $\tau = 0.75$, the mean lengths of the confidence interval for $\hat{\beta}_2(\tau)$ are similar for both methods, at 0.20 and 0.22, respectively. However, our method (OM) has a lower standard error (the bracketed number) of 0.028 compared to 0.034 for the NPKE method. Notably, in this example, both methods have a similar empirical coverage probability around 95% (See Table 3.5).

(2^0) For quantiles $\tau = 0.1$ and 0.9, the proposed approach (OM) exhibits wider confidence intervals than the NPKE approach but with similar standard errors for the mean length. This behavior can be explained by the pattern observed in Tables 3.3 and 3.4 for the estimation of the limiting variance $\Gamma(\hat{\beta}_j(\tau))$ when using $\tilde{f}_Z$. This is, for

Table 3.5. Comparison of empirical coverage rates of the 95% confidence intervals for the estimated coefficients $\hat{\beta}_0(\tau)$, $\hat{\beta}_1(\tau)$ and $\hat{\beta}_2(\tau)$ under the three models considered at different sample sizes n and quantile levels τ.

			$\hat{\beta}_0(\tau)$		$\hat{\beta}_1(\tau)$		$\hat{\beta}_2(\tau)$	
	n	τ	OM	NKDE	OM	NKDE	OM	NKDE
Model 1	100	0.10	0.91	0.88	0.93	0.90	0.93	0.90
		0.25	0.94	0.96	0.95	0.97	0.94	0.96
		0.50	0.95	0.98	0.94	0.97	0.96	0.98
		0.75	0.94	0.96	0.95	0.96	0.94	0.96
		0.90	0.90	0.87	0.94	0.91	0.93	0.90
	400	0.10	0.93	0.91	0.93	0.91	0.94	0.92
		0.25	0.94	0.96	0.94	0.96	0.94	0.96
		0.50	0.95	0.97	0.96	0.98	0.95	0.97
		0.75	0.94	0.96	0.95	0.96	0.94	0.95
		0.90	0.94	0.93	0.95	0.93	0.94	0.92
	1000	0.10	0.94	0.93	0.94	0.92	0.94	0.93
		0.25	0.95	0.96	0.95	0.96	0.95	0.97
		0.50	0.94	0.96	0.95	0.97	0.96	0.97
		0.75	0.96	0.97	0.95	0.96	0.95	0.96
		0.90	0.93	0.91	0.94	0.93	0.94	0.93
Model 2	100	0.10	0.90	0.86	0.92	0.89	0.90	0.87
		0.25	0.94	0.96	0.95	0.96	0.95	0.96
		0.50	0.94	0.98	0.96	0.98	0.94	0.97
		0.75	0.94	0.97	0.94	0.96	0.95	0.97
		0.90	0.91	0.87	0.91	0.87	0.92	0.89
	400	0.10	0.93	0.91	0.94	0.92	0.94	0.92
		0.25	0.94	0.96	0.94	0.96	0.94	0.95
		0.50	0.95	0.98	0.96	0.98	0.95	0.97
		0.75	0.94	0.96	0.94	0.96	0.95	0.96
		0.90	0.91	0.88	0.92	0.90	0.93	0.91
	1000	0.10	0.93	0.91	0.94	0.91	0.95	0.94
		0.25	0.94	0.95	0.96	0.97	0.95	0.96
		0.50	0.95	0.97	0.96	0.97	0.95	0.97
		0.75	0.95	0.96	0.93	0.95	0.96	0.97
		0.90	0.93	0.91	0.93	0.91	0.93	0.92
Model 3	100	0.10	0.92	0.89	0.93	0.89	0.93	0.91
		0.25	0.95	0.96	0.94	0.97	0.95	0.97
		0.50	0.95	0.98	0.95	0.98	0.95	0.98
		0.75	0.97	0.98	0.95	0.98	0.95	0.96
		0.90	0.94	0.91	0.94	0.91	0.94	0.91
	400	0.10	0.95	0.93	0.93	0.91	0.94	0.93
		0.25	0.96	0.97	0.95	0.96	0.96	0.97
		0.50	0.95	0.97	0.94	0.97	0.94	0.96
		0.75	0.95	0.97	0.95	0.96	0.96	0.97
		0.90	0.94	0.91	0.93	0.92	0.95	0.93
	1000	0.10	0.94	0.93	0.94	0.93	0.94	0.93
		0.25	0.95	0.96	0.95	0.97	0.96	0.97
		0.50	0.94	0.96	0.97	0.98	0.96	0.97
		0.75	0.96	0.97	0.95	0.96	0.95	0.97
		0.90	0.93	0.92	0.95	0.93	0.95	0.94

Table 3.6. Comparison of the average length of nominal 95% confidence intervals for the estimated coefficients $\hat{\beta}_0(\tau)$, $\hat{\beta}_1(\tau)$ and $\hat{\beta}_2(\tau)$ under the three models considered at different sample sizes n and quantile levels τ. The bracketed number gives the standard error of each average length.

			$\hat{\beta}_0(\tau)$		$\hat{\beta}_1(\tau)$		$\hat{\beta}_2(\tau)$	
	n	τ	OM	NKDE	OM	NKDE	OM	NKDE
Model 1	100	0.10	0.75 (0.160)	0.67 (0.132)	0.11 (0.024)	0.10 (0.020)	0.26 (0.055)	0.23 (0.045)
		0.25	0.56 (0.079)	0.63 (0.095)	0.08 (0.013)	0.09 (0.015)	0.19 (0.026)	0.22 (0.031)
		0.50	0.51 (0.068)	0.62 (0.094)	0.08 (0.011)	0.09 (0.014)	0.18 (0.022)	0.22 (0.030)
		0.75	0.56 (0.087)	0.63 (0.102)	0.08 (0.013)	0.10 (0.015)	0.20 (0.028)	0.22 (0.034)
		0.90	0.75 (0.159)	0.67 (0.129)	0.11 (0.024)	0.10 (0.020)	0.26 (0.054)	0.23 (0.044)
	400	0.10	0.38 (0.046)	0.35 (0.042)	0.06 (0.007)	0.05 (0.006)	0.13 (0.016)	0.12 (0.014)
		0.25	0.28 (0.021)	0.30 (0.024)	0.04 (0.003)	0.05 (0.004)	0.10 (0.007)	0.10 (0.008)
		0.50	0.25 (0.016)	0.29 (0.021)	0.04 (0.002)	0.04 (0.003)	0.09 (0.005)	0.10 (0.007)
		0.75	0.28 (0.020)	0.30 (0.024)	0.04 (0.003)	0.05 (0.004)	0.10 (0.007)	0.10 (0.008)
		0.90	0.38 (0.042)	0.35 (0.039)	0.06 (0.006)	0.05 (0.006)	0.13 (0.014)	0.12 (0.013)
	1000	0.10	0.24 (0.017)	0.23 (0.017)	0.04 (0.003)	0.03 (0.003)	0.08 (0.006)	0.08 (0.006)
		0.25	0.18 (0.008)	0.19 (0.010)	0.03 (0.001)	0.03 (0.001)	0.06 (0.003)	0.06 (0.003)
		0.50	0.16 (0.006)	0.18 (0.008)	0.02 (0.001)	0.03 (0.001)	0.06 (0.002)	0.06 (0.003)
		0.75	0.18 (0.008)	0.19 (0.009)	0.03 (0.001)	0.03 (0.001)	0.06 (0.003)	0.07 (0.003)
		0.90	0.24 (0.018)	0.23 (0.018)	0.04 (0.003)	0.03 (0.003)	0.08 (0.006)	0.08 (0.006)
Model 2	100	0.10	0.30 (0.069)	0.27 (0.057)	0.05 (0.011)	0.04 (0.009)	0.10 (0.023)	0.09 (0.019)
		0.25	0.22 (0.035)	0.25 (0.041)	0.03 (0.005)	0.04 (0.006)	0.08 (0.011)	0.09 (0.014)
		0.50	0.20 (0.028)	0.24 (0.037)	0.03 (0.004)	0.04 (0.006)	0.07 (0.009)	0.08 (0.012)
		0.75	0.22 (0.034)	0.25 (0.040)	0.03 (0.005)	0.04 (0.006)	0.08 (0.011)	0.09 (0.013)
		0.90	0.30 (0.065)	0.27 (0.053)	0.05 (0.010)	0.04 (0.008)	0.10 (0.022)	0.09 (0.018)
	400	0.10	0.15 (0.019)	0.14 (0.018)	0.02 (0.003)	0.02 (0.003)	0.05 (0.006)	0.05 (0.006)
		0.25	0.11 (0.009)	0.12 (0.010)	0.02 (0.001)	0.02 (0.001)	0.04 (0.003)	0.04 (0.003)
		0.50	0.10 (0.007)	0.11 (0.009)	0.01 (0.001)	0.02 (0.001)	0.03 (0.002)	0.04 (0.003)
		0.75	0.11 (0.009)	0.12 (0.010)	0.02 (0.001)	0.02 (0.001)	0.04 (0.003)	0.04 (0.003)
		0.90	0.15 (0.018)	0.14 (0.018)	0.02 (0.003)	0.02 (0.003)	0.05 (0.006)	0.05 (0.006)
	1000	0.10	0.10 (0.008)	0.09 (0.007)	0.01 (0.001)	0.01 (0.001)	0.03 (0.003)	0.03 (0.003)
		0.25	0.07 (0.003)	0.07 (0.004)	0.01 (0.001)	0.01 (0.001)	0.02 (0.001)	0.03 (0.001)
		0.50	0.06 (0.003)	0.07 (0.003)	0.01 (0.001)	0.01 (0.001)	0.02 (0.001)	0.02 (0.001)
		0.75	0.07 (0.003)	0.07 (0.004)	0.01 (0.001)	0.01 (0.001)	0.02 (0.001)	0.03 (0.001)
		0.90	0.10 (0.008)	0.09 (0.008)	0.01 (0.001)	0.01 (0.001)	0.03 (0.003)	0.03 (0.003)
Model 3	100	0.10	0.67 (0.120)	0.61 (0.103)	0.10 (0.018)	0.09 (0.015)	0.23 (0.040)	0.21 (0.034)
		0.25	0.52 (0.060)	0.58 (0.070)	0.08 (0.009)	0.09 (0.011)	0.18 (0.018)	0.20 (0.022)
		0.50	0.48 (0.048)	0.57 (0.067)	0.07 (0.008)	0.09 (0.010)	0.17 (0.016)	0.20 (0.022)
		0.75	0.52 (0.064)	0.58 (0.074)	0.08 (0.010)	0.09 (0.011)	0.18 (0.019)	0.20 (0.023)
		0.90	0.67 (0.122)	0.60 (0.099)	0.10 (0.019)	0.09 (0.015)	0.23 (0.040)	0.21 (0.032)
	400	0.10	0.33 (0.030)	0.31 (0.030)	0.05 (0.005)	0.05 (0.005)	0.12 (0.010)	0.11 (0.010)
		0.25	0.26 (0.015)	0.28 (0.018)	0.04 (0.002)	0.04 (0.003)	0.09 (0.005)	0.10 (0.006)
		0.50	0.24 (0.012)	0.26 (0.016)	0.04 (0.002)	0.04 (0.002)	0.08 (0.004)	0.09 (0.005)
		0.75	0.26 (0.015)	0.28 (0.018)	0.04 (0.002)	0.04 (0.003)	0.09 (0.005)	0.10 (0.006)
		0.90	0.33 (0.030)	0.31 (0.030)	0.05 (0.005)	0.05 (0.005)	0.12 (0.010)	0.11 (0.010)
	1000	0.10	0.21 (0.012)	0.20 (0.013)	0.03 (0.002)	0.03 (0.002)	0.07 (0.004)	0.07 (0.004)
		0.25	0.16 (0.006)	0.17 (0.008)	0.02 (0.001)	0.03 (0.001)	0.06 (0.002)	0.06 (0.003)
		0.50	0.15 (0.005)	0.16 (0.006)	0.02 (0.001)	0.02 (0.001)	0.05 (0.002)	0.06 (0.002)
		0.75	0.16 (0.006)	0.17 (0.007)	0.02 (0.001)	0.03 (0.001)	0.06 (0.002)	0.06 (0.002)
		0.90	0.21 (0.012)	0.20 (0.013)	0.03 (0.002)	0.03 (0.002)	0.07 (0.004)	0.07 (0.005)

these values of τ, we have $\hat{\Gamma}_{\text{NKDE}}(\hat{\beta}_j(\tau)) < \hat{\Gamma}(\hat{\beta}_j(\tau))$, resulting in wider confidence intervals when using the proposed estimator in the simulation study.

(3^0) As n increases, both the average lengths and their respective standard errors decrease.

Overall, based on the results from Table 3.5 and Table 3.6, we can conclude that it is better to use our method (OM) compared to the NKDE method, to estimate the limiting variance of $\hat{\beta}_j(\tau)$'s , as it leads to better performance of the confidence interval for the estimated coefficients in the quantile regression model.

3.3.4 Quantile regression application to the monthly inflation rate for the United States

The behavior of prices, in some way, the change of prices (inflation rate), is a key variable in the economy, which determines the population's purchasing power and affects the investment decisions of different agents. Due to its relevance, institutions such as central banks, with scheme inflation target, try to keep the inflation rate low and stable in order to preserve the purchasing power of money and thus contribute to the economy producing at the maximum level of its capacities. Therefore, this is one of the most relevant economic series, especially for countries with target inflation, which follows the changes (variations) in the price of a basket of goods and services representative of the consumption of households (See IMF (2004)). The percentage change of the consumer price index (CPI) is a typically used measure of the inflation rate, which is a statistical measure, over time, of the prices of goods and services in major expenditure groups. Thus, in this Subsection, to apply the proposed density estimate in the context of quantile regression with nonparametric ARCH errors, we analyze the monthly inflation time series of the total CPI of the United States.

We consider the monthly not seasonally adjusted total CPI (1982–1984 = 100) of the United States, available at the U.S. Bureau of Labor Statistics web (`https://www.bls.gov/cpi/data.htm`). The data consists of $n = 241$ monthly observations for the time series from September 2003 to September 2023. This data was previously analyzed by various authors, for example, by Li and Zhao (2019), in the case of two monthly CPI measures: the seasonally adjusted total CPI and the core CPI, which

excludes food and energy. The monthly inflation rate Y_i is the percentage change of CPI from month i to $i-1$:

$$Y_i = \frac{\text{CPI}_i}{\text{CPI}_{i-1}} - 1,$$

where CPI_i is the CPI at month i.

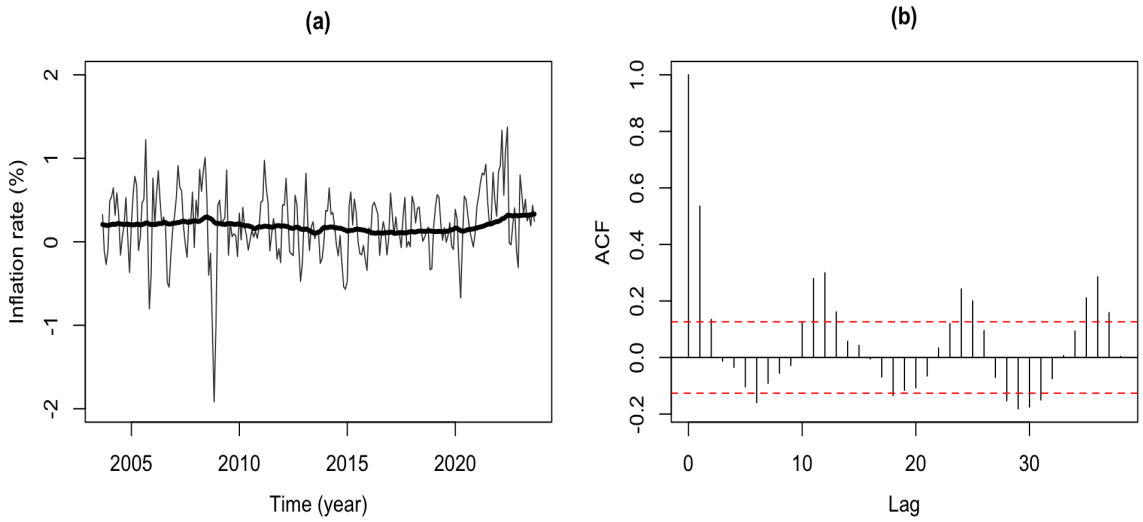

Figure 3.5. (a) Monthly total inflation rate of the United States, and 60-month moving average trend (▬) during September 2003-September 2023, and (b) The estimated sample autocorrelation function for the monthly total inflation rate.

Figure 3.5(a) displays the monthly total inflation rate Y_i. The graph shows that the volatility in the dataset varies over time, with some periods being more volatile than others. Specifically, the period around 2008 exhibits significantly higher volatility than any other time frame considered in this study. However, after the 2008 crisis, the monthly total inflation rate appeared to exhibit less volatility, which is similar to the earlier 2004. Similar to Li and Zhao (2019), we calculate the 60-month moving average trend (the thick full curve in Figure 3.5(a)) to visualize the general behavior of the data set. Thus, this 60-month moving average trend shows that the time series appears stationary around a constant mean. In fact, as the stationarity test of Kwiatkowski et al. (1992) (`R` command `kpss.test` in package `tseries`) has a p-value of 0.1, the null hypothesis of stationarity is not rejected at a level of significance 0.05. Also, the unit root test in Perron (1988) (`R` command `pp.test`) indicates that the time series does not

have a unit root (p-value=0.01), and then it is stationary. Thus, based on the observed behavior of changing time series variability and the fact that the time series is stationary, it is possible to suggest that an appropriate model for this case should have a volatility structure that varies with time, such as the autoregressive model with ARCH errors (AR-ARCH).

To visualize the serial correlation and the no presence of a unit root in the monthly total inflation rate, we use Figure 3.5(b). The graph exhibits the estimated sample autocorrelation function (ACF), where the dashed red lines show the 95% confidence interval for these estimated sample ACF. The ACF shows a fast decaying of the autocorrelations, which confirms that the time series does not have a unit root. Also, the seasonal autocorrelation relationships are shown quite prominently, as there are significant correlations at lags 12, 24, 36, and so on, which suggests the possibility of including a seasonal component in the model.

We use the traditional time series approach to find an adequate model for our variable of interest, as we are more interested in estimating the quantity $f_Z(F_Z^{-1}(\tau))$ present in the asymptotic variance of $\hat{\beta}(\tau)$ than in finding the best model for the inflation rate to forecast. Thus, using the Akaike information criterion, we select lags 1, 2, and 12 of the monthly total inflation rate as the explanatory variables that generate an appropriate model. For $1 \leq i \leq n$, we suggest that the following autoregressive regression model with nonparametric ARCH errors can be an adequate model for the monthly total inflation rate data:

$$Y_i = \beta_0 + \beta_1 Y_{i-1} + \beta_2 Y_{i-2} + \beta_3 Y_{i-12} + Z_i \quad \text{with} \quad Z_i \sim \text{ARCH model in (3.1)}. \quad (3.29)$$

Having specified the monthly total inflation rate model, we present estimation results for the vector $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^T$ using the quantile regression estimator in (3.18) (`R` command `qr` in package `quantreg`) together with their 95% confidence intervals using the two methods described in Subsection 3.3.3. To compute the confidence interval, for any fixed τ, of each $\hat{\beta}_j(\tau)$, $j = 0, 1, 2, 3$, we need to estimate its limiting variance, $\hat{\Gamma}(\hat{\beta}_j(\tau))$ which depends on the reciprocal of the density function of the model error terms, $1/f_Z^2(F_Z^{-1}(\tau))$. For comparison, we estimate $f_Z(F_Z^{-1}(\tau))$ using the two methods: OM method and NKDE method.

In the first method (OM), for fixed value $\tau \in (0, 1)$, based on the centered estimated residuals of the model (3.29), we estimate $f_Z(F_Z^{-1}(\tau))$ using the proposed density estimator

$\hat{f}_{\hat{Z}}(\hat{F}_{\hat{Z}}^{-1}(\tau))$ in (3.5) and thus we get the estimation of the limiting variance $\hat{\Gamma}(\hat{\beta}_j(\tau))$, for $j = 0, 1, 2, 3$. In the second method (NKDE), we use the nonparametric kernel density estimator $\tilde{f}_{\hat{Z}}(\hat{F}_{\hat{Z}}^{-1}(\tau))$ in (3.2) to estimate the quantity $f_Z(F_Z^{-1}(\tau))$ and in the same way get an estimation of the limiting variance.

Table 3.7 presents the length of the estimated 95% confidence intervals (C.I.) for the estimated coefficients $\hat{\beta}_0(\tau)$, $\hat{\beta}_1(\tau)$, $\hat{\beta}_2(\tau)$ and $\hat{\beta}_3(\tau)$ in the model (3.29), using the two methods: the proposed estimator $\hat{f}_{\hat{Z}}$ (OM method) and the nonparametric kernel density estimator $\tilde{f}_{\hat{Z}}$ (NKDE method). Seven quantiles were considered: 0.05, 0.10, 0.25, 0.50, 0.75, 0.90 and 0.95. It can be seen that for quantiles τ close to 0.5, specifically for τ=0.25, 0.50, and 0.75, both methods show similar confidence interval lengths. However, for most cases, our method (OM) produces slightly shorter lengths of confidence intervals than the NKDE method. For values of τ less than 0.1 and greater than 0.9, the OM method produces wider confidence intervals. As mentioned in the last Subsection 3.3.3, the behavior in the length of C.I. associated with small and high levels of τ is due to the positive bias in the estimator $\tilde{f}_{\hat{Z}}$ (NKDE method). To illustrate the behavior of the two methods in estimating the confidence intervals of the estimated coefficients in more detail, Figure (3.6) shows the plots of these confidence intervals for a sequence of different quantiles τ.

Table 3.7. Comparison of length of the estimated 95% confidence intervals for the estimated coefficients $\hat{\beta}_0(\tau)$, $\hat{\beta}_1(\tau)$, $\hat{\beta}_2(\tau)$ and $\hat{\beta}_3(\tau)$ in the model (3.29), by utilizing the proposed estimator $\hat{f}_{\hat{Z}}$ (OM method) and the usual nonparametric kernel density estimator $\tilde{f}_{\hat{Z}}$ (NKDE method) for seven quantiles τ.

	Length C.I of $\hat{\beta}_0(\tau)$		Length C.I of $\hat{\beta}_1(\tau)$		Length C.I of $\hat{\beta}_2(\tau)$		Length C.I of $\hat{\beta}_3(\tau)$	
τ	OM	NKDE	OM	NKDE	OM	NKDE	OM	NKDE
0.05	0.19	0.16	0.47	0.40	0.45	0.39	0.40	0.34
0.10	0.17	0.16	0.41	0.40	0.40	0.38	0.35	0.34
0.25	0.13	0.13	0.32	0.33	0.31	0.32	0.27	0.28
0.50	0.12	0.13	0.31	0.32	0.30	0.31	0.26	0.27
0.75	0.13	0.14	0.32	0.35	0.31	0.33	0.27	0.29
0.90	0.18	0.17	0.45	0.43	0.44	0.41	0.38	0.36
0.95	0.20	0.17	0.48	0.41	0.47	0.39	0.41	0.35

Figure 3.6 summarizes the quantile regression results for the application to the monthly total inflation rate. Each plot displays one estimated coefficient from the quantile autoregressive model with nonparametric ARCH errors in (3.29). We consider a sequence of 100 equally spaced quantile levels τ between 0.05 and 0.95. The black dots

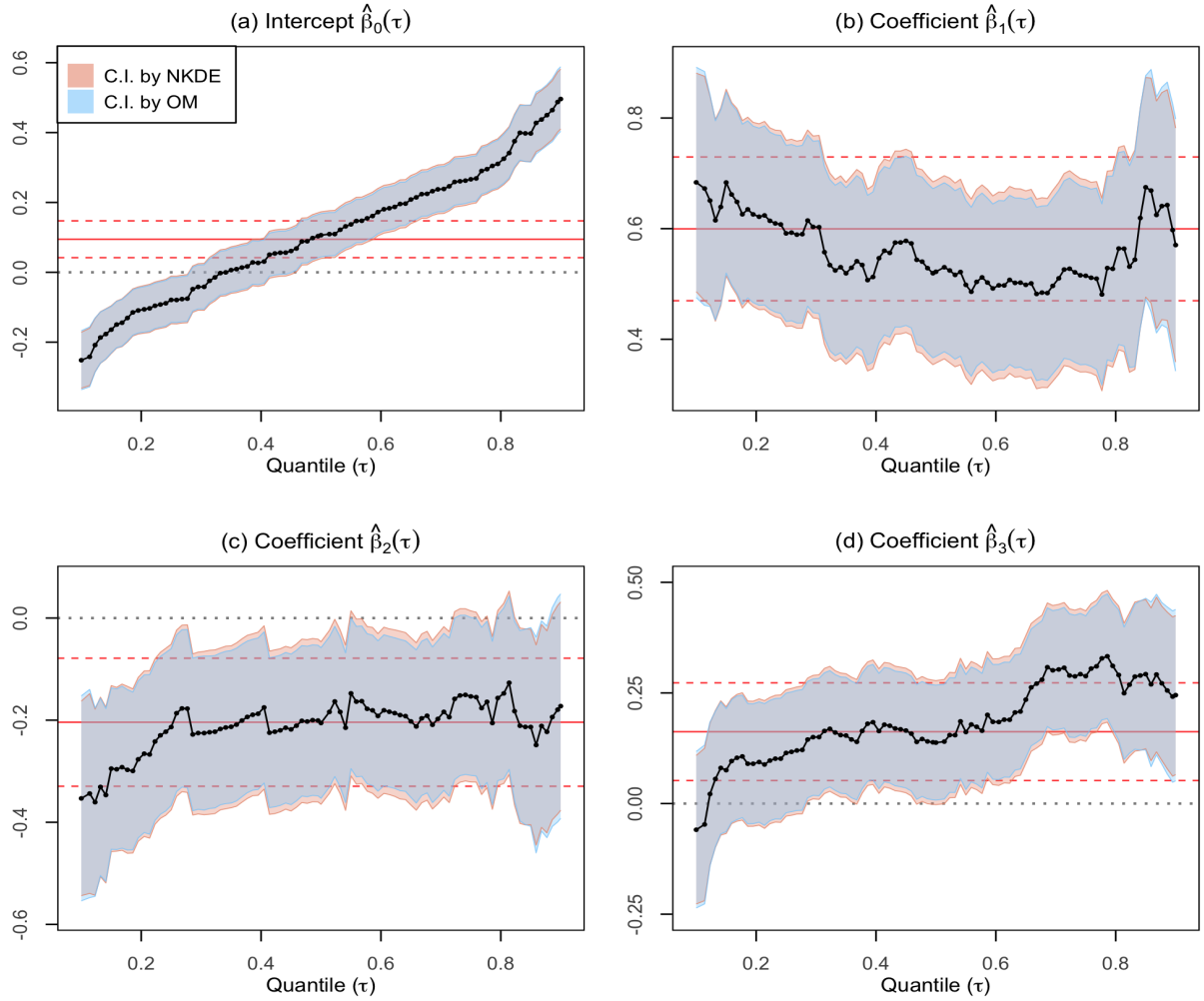

Figure 3.6. Estimated 95% confidence intervals (C.I) for $\hat{\beta}(\tau)$ in Model (3.29) using two methods to compute the asymptotic variance: the first method, using $\tilde{f}_{\hat{Z}}$ (NKDE method), is indicated by the shaded area marked in salmon color ( ), and the second method, the proposed method (OM) using $\hat{f}_{\hat{Z}}$, is indicated by the shaded area in blue ( ). The red line on each plot represents the OLS estimate of the mean effect, with two red dashed lines again representing a 95% C.I for this coefficient. The gray dotted line $(\cdots\cdot)$ indicates $\beta = 0$, the null effect. (a) C.I for intercept $\hat{\beta}_0(\cdot)$; (b) C.I for coefficient $\hat{\beta}_1(\cdot)$; (c) C.I for coefficient $\hat{\beta}_2(\cdot)$; (d) C.I for coefficient $\hat{\beta}_3(\cdot)$.

represent the estimated coefficients, $\{\hat{\beta}_j(\tau) : j = 0, 1, 2, 3.\}$, for the quantiles indicated on the x-axis, with a solid black line connecting them. The shaded area marked in blue represents the estimated 95% confidence intervals using the proposed method (OM), while the shaded area marked in salmon color represents the estimated 95% confidence intervals using the NKDE method. Additionally, a red line is superimposed on each plot, representing the common ordinary least squares (OLS) estimate of the mean effect, with two red dashed lines again representing the OLS 95% confidence interval for this

coefficient. The gray dotted line $(\cdots\cdot)$ indicates $\beta_j = 0$, the null effect. We corroborate here the fact that for quantiles close to 0.5 the proposed method (OM) results in narrower confidence intervals than the NKDE method for $\hat{\beta}_1(\cdot)$, $\hat{\beta}_2(\cdot)$ and $\hat{\beta}_3(\cdot)$. In contrast, for small and high quantiles, our method (OM) exhibits wider confidence intervals compared to the NKDE method. In this case, it is better to use our proposed density estimator $\hat{f}_{\hat{Z}}(z)$ in (3.5) (OM method), based on the estimated residuals of the model, to estimate the asymptotic variance of $\hat{\beta}_j(\tau)$'s in the quantile regression model compared with the NKDE method.

3.4 Proofs

Write $\|Z\|^2 = \mathbb{E}(Z^2)$ for a random variable Z. Throughout, denote by $\phi(z;\sigma^2)$ the normal density of $N(0,\sigma^2)$, and denote the partial derivative by $\dot{\phi}_{\sigma^2}(z;\sigma^2) = \partial\phi(z;\sigma^2)/\partial\sigma^2$. Also, $c_1, c_2, \ldots,$ are generic constants that may vary from places to places, and $\mathcal{Z}$ is any bounded interval with length denoted by $|\mathcal{Z}|$. For convenience, throughout we write $K_{h_n}(u) = K(u/h_n)/h_n$.

Lemma 5 below is from Section 2.4 in Chapter 2, and the proof is omitted here.

Lemma 5. *For any differentiable random function $H(y)$,*

$$\sup_{y\in\mathcal{Y}} |H(y)| = O_p\left\{\inf_{y\in\mathcal{Y}} \|H(y)\| + \sup_{y\in\mathcal{Y}} \|H'(y)\|\right\}. \tag{3.30}$$

Lemma 6. *Let $g(\cdot;\cdot)$ and $h(\cdot)$ be functions such that $\mathbb{E}h(e_0) = 0$, $\mathbb{E}h^2(e_0) < \infty$, $\sup_{x\in\mathcal{X}} \mathbb{E}[g^2(z;Z_0) + \dot{g}^2(z;Z_0)] < c$. Define*

$$\Lambda_n(z) = \sum_{j=1}^{n} g(z;Z_{j-1})h(e_j).$$

Then $\sup_{z\in\mathcal{Z}} |\Lambda_n(z)| = O_p(n^{1/2})$.

Proof. Recall the sigma filed $\mathcal{F}_i$ in Assumption 6. By causality, $\mathbb{E}[g(z;X_{j-1})h(e_j)|\mathcal{F}_{j-1}] = g(z;Z_{j-1})\mathbb{E}[h(e_j)|\mathcal{F}_{j-1}] = 0$ via the i.i.d. assumption of $\{e_j\}_j$. Thus, for each fixed z,

$\{g(z;Z_{j-1})h(e_j)\}_i$ form martingale differences. By orthogonality of martingale differences,

$$\begin{aligned}\mathbb{E}\Lambda_n(z)^2 = \sum_{j=1}^{n}\mathbb{E}[g^2(z;Z_{j-1})h^2(e_j)] &= \sum_{j=1}^{n}\mathbb{E}g^2(z;Z_{j-1})\mathbb{E}h^2(e_j)\\ &\leq n\mathbb{E}h^2(e_0)\sup_{z\in\mathcal{Z}}\mathbb{E}[g^2(z;Z_0).\end{aligned}$$

Similarly, for the derivative $\Lambda_n'(z) = \sum_{i=1}^{n}\dot{g}(z;Z_{j-1})h(e_j)$, by the same argument, we have

$$\mathbb{E}\Lambda_n'(z)^2 \leq n\mathbb{E}h^2(e_0)\sup_{z\in\mathcal{Z}}\mathbb{E}[\dot{g}^2(z;Z_0).$$

The desired result then follows from Lemma 5 under the specified condition. ♢

Lemma 7. *Let function $g(\cdot;\cdot)$ satisfy $\sup_{z\in\mathcal{Z}}\mathbb{E}[g^2(z;Z_0)+\dot{g}^2(z;Z_0)] < c$. Define*

$$\Xi_{j,n}(z) = \frac{1}{n}\sum_{i=1}^{n}K_{h_n}(Z_{i-1}-Z_{j-1})g(z;Z_{i-1}).$$

Then

$$\sup_{1\leq j\leq n, z\in\mathcal{Z}}|\Xi_{j,n}(z) - f_Z(Z_{j-1})g(z;Z_{j-1})| = O_p\left(h_n^2 + \frac{\log n}{\sqrt{nh_n}}\right).$$

Proof. Define

$$\Xi_n(x,z) = \frac{1}{n}\sum_{i=1}^{n}K_{h_n}(Z_{i-1}-x)g(z;Z_{i-1}).$$

For example, the special case of $g(\cdot;\cdot)=1$ is just the nonparametric kernel density estimate. Following the same lines of proofs in Hansen (2008), we can show that the following approximation holds

$$\Xi_n(x,z) = f_Z(z)g(z;x) + O_n\left(h_n^2 + \frac{\log n}{\sqrt{nh_n}}\right), \quad \text{uniformly over } z\in\mathcal{Z},\ |x| = O(n). \quad (3.31)$$

On the other hand, from the condition $\mathbb{E}|Z_0|^s < \infty$ for some $s>2$, we have $\max_j|Z_j| = O_p(\sqrt{n})$, which together with (3.31) gives the desired result. ♢

Proof of Theorem 6. Under Assumptions 5 and 6, by Hansen (2008), we have

$$\sup_{|z|=O(n)} \left|\hat{\sigma}^2(z) - \sigma^2(z)\right| = O_p\left(h_n^2 + \frac{\log n}{\sqrt{nh_n}}\right) \tag{3.32}$$

The expression in (3.32) provides a uniform error bound for the estimate $\hat{\sigma}^2(\cdot)$. Define

$$\delta_n := \sup_{1\le i\le n} \left|\hat{\sigma}^2\left(Z_{i-1}\right) - \sigma^2\left(Z_{i-1}\right)\right|^2. \tag{3.33}$$

As mentioned in the proof of Lemma 7, we have $\max_i |Z_i| = O_p(\sqrt{n})$, which together with (3.32) implies that

$$\delta_n = O_p\left(h_n^4 + \frac{(\log n)^2}{nh_n}\right) = o_p(n^{-1/2}), \quad \text{under bandwidth condition in Assumption 5.} \tag{3.34}$$

Note that, the first and second order derivatives of $\phi\left(z;\sigma^2\right)$ with respect to $\sigma^2 > \sigma_0^2$ are all uniformly bounded in $z \in \mathbb{R}$. Applying Taylor's expansion of $\phi\left(z;\hat{\sigma}^2\left(Z_{i-1}\right)\right)$ at $\sigma^2\left(Z_{i-1}\right)$ and using (3.33) and (3.34), we can obtain

$$\hat{f}_Z(z) = \frac{1}{n}\sum_{i=1}^{n} \phi\left(z;\sigma^2\left(Z_{i-1}\right)\right) + I_n(z) + o_p(n^{-1/2}), \tag{3.35}$$

where

$$I_n(z) = \frac{1}{n}\sum_{i=1}^{n} \left[\hat{\sigma}^2\left(Z_{i-1}\right) - \sigma^2\left(Z_{i-1}\right)\right] \dot{\phi}\left(z;\sigma^2\left(Z_{i-1}\right)\right) \quad \text{with} \quad \dot{\phi}\left(z;\sigma^2\right) = \frac{\partial\phi(z;\sigma^2)}{\partial\sigma^2}. \tag{3.36}$$

For the nonparametric estimate $\hat{\sigma}^2(z)$ in (3.6), recall the notation $K_{h_n}(u) = K\left(u/h_n\right)/h_n$, we can rewrite

$$\hat{\sigma}^2(z) = \frac{\sum_{j=1}^{n} K_{h_n}\left(z - Z_{j-1}\right) Z_j^2}{n\tilde{f}_Z(z)} \quad \text{where} \quad \tilde{f}_Z(z) = \frac{1}{n}\sum_{j=1}^{n} K_{h_n}\left(z - Z_{j-1}\right). \tag{3.37}$$

Using $Z_j = \sigma\left(Z_{j-1}\right) e_j$, we can obtain

$$\hat{\sigma}^2(z) - \sigma^2(z) = B_n(z) + \frac{\sum_{j=1}^{n} K_{h_n}\left(z - Z_{j-1}\right)\sigma^2(z)\left[e_j^2 - 1\right]}{n\tilde{f}_Z(z)} \tag{3.38}$$

with

$$B_n(z) = \frac{\sum_{j=1}^{n} K_{h_n}\left(z - Z_{j-1}\right)\left[\sigma^2\left(Z_{j-1}\right) - \sigma^2(z)\right] e_j^2}{n\tilde{f}_Z(z)}. \tag{3.39}$$

The bias term $B_n(z)$ satisfies $\sup_z |B_n(z)| = O_p(h_n^2)$. Thus, by (3.38) and the uniform boundedness of $\dot{\phi}\left(z; \sigma\left(Z_{i-1}\right)\right)$, we have:

$$I_n(z) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{K_{h_n}\left(Z_{i-1} - Z_{j-1}\right)\sigma^2(Z_{i-1})\left[e_j^2 - 1\right]}{n\tilde{f}_Z(Z_{i-1})}\dot{\phi}\left(z; \sigma^2\left(Z_{i-1}\right)\right) + O_p\left(h_n^2\right). \tag{3.40}$$

For $\tilde{f}_Z(z)$ in (3.37), by property [Hansen (2008)] of nonparametric kernel density estimator,

$$\tilde{f}_Z(z) = f_Z(z) + O_p\left(h_n^2 + \sqrt{\frac{\log n}{nh_n}}\right), \quad \text{uniformly in } z. \tag{3.41}$$

Here $f_Z(z)$ is the density of Z_i. Substituting this in (3.40), we can express $I_n(z)$ as

$$I_n(z) = \left[1 + O_p\left(h_n^2 + \sqrt{\frac{\log n}{nh_n}}\right)\right]\tilde{I}_n(z) + O_p\left(h_n^2\right), \tag{3.42}$$

where

$$\tilde{I}_n(z) = \frac{1}{n}\sum_{j=1}^{n}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{K_{h_n}\left(Z_{i-1} - Z_{j-1}\right)\dot{\phi}\left(z; \sigma^2\left(Z_{i-1}\right)\right)\sigma^2(Z_{i-1})}{f_Z\left(Z_{i-1}\right)}\right](e_j^2 - 1).$$

By Lemma 7 and the fact $\sum_{j=1}^{n}(e_j^2 - 1) = O_p(\sqrt{n})$, we can further approximate $\tilde{I}_n(z)$ as

$$\tilde{I}_n(z) = \tilde{I}_n^*(z) + O_p\left(h_n^2 + \frac{\log n}{n\sqrt{h_n}}\right), \quad \text{where} \quad \tilde{I}_n^*(z) = \frac{1}{n}\sum_{j=1}^{n}\dot{\phi}\left(z; \sigma^2\left(Z_{j-1}\right)\right)\sigma^2(Z_{j-1})(e_j^2 - 1). \tag{3.43}$$

By Lemma 6, $\sup_{z\in\mathcal{Z}} |\tilde{I}_n^*(z)| = O_p(n^{-1/2})$. Combining this latter fact with (3.42), (3.43), and the bandwidth condition in Assumption 5, we have

$$I_n(z) = \frac{1}{n}\sum_{j=1}^{n}\dot{\phi}\left(z; \sigma^2\left(Z_{j-1}\right)\right)\sigma^2(Z_{j-1})(e_j^2 - 1) + o_p(n^{-1/2}). \tag{3.44}$$

From (3.35), (3.44) and the fact $f_Z(z) = \mathbb{E}\phi(z; \sigma^2(Z_{i-1}))$, we can obtain

$$\hat{f}_Z(z) - f_Z(z) = \frac{1}{n}\sum_{i=1}^{n} \xi_i(z) + o_p(n^{-1/2}), \tag{3.45}$$

where $\xi_i(z)$ is defined as in (3.9). It is well known that central limit theorems hold under the mixing condition in Assumption 6. The finite conditional convergence then follows from the Cramér-Wold device, and the tightness clearly holds from the differentiability of $\xi_i(z)$. This completes the proof. ♢

Proof of Proposition 1. From the proof of Theorem 6, $\hat{\sigma}^2(Z_{i-1}) = \sigma^2(Z_{i-1}) + o_p(1)$, uniformly over i. Thus, from the uniform consistency of $\hat{f}_Z(z)$ and the boundedness of $\phi(z; \sigma^2)$ and $\dot{\phi}(z; \sigma^2)$, we can easily obtain

$$\begin{aligned}\hat{\xi}_i(z) &= [\phi(z; \sigma^2(Z_{i-1}) - f_Z(z) + o_p(1)] + [\dot{\phi}(z; \sigma^2(Z_{i-1})) + o_p(1)][Z_i^2 - \sigma^2(Z_{i-1}) + o_p(1)] \\ &= \xi_i(z) + o_p(1) + o_p(1)[Z_i^2 + \sigma^2(Z_{i-1})],\end{aligned}$$

uniformly over $1 \le i \le n$ and $z \in \mathcal{Z}$. The result then easily follows from the law of large numbers. ♢

Proof of Corollary 2. In (3.14), for Gaussian kernel $K(u)$, $\int_{\mathbb{R}} u^2 K(u)du = 1$. Let $Z = (Z_1, ..., Z_{i-1}, ..., Z_{n-1})$. Now, to compute $f_Z''(z)$, under the nonparametric autoregressive conditional heteroscedasticity model in (3.1), note that by (3.3) the density function f_Z can be written as

$$f_Z(z) = \mathbb{E}\left[\frac{1}{\sigma(Z)}\phi\left(\frac{z}{\sigma(Z)}; 1\right)\right], \tag{3.46}$$

where $\phi(\cdot; 1)$ denotes the standard normal density of $N(0, 1)$. Thus,

$$f_Z''(z) = \mathbb{E}\left[\frac{1}{\sigma(Z)}\frac{\partial^2}{\partial z^2}\phi\left(\frac{z}{\sigma(Z)}; 1\right)\right], \tag{3.47}$$

Elementary calculations show

$$\frac{\partial^2}{\partial z^2}\phi\left(\frac{z}{\sigma(Z)}; 1\right) = \frac{1}{\sigma^2(Z)}\left(\frac{z^2}{\sigma^2(Z)} - 1\right)\phi\left(\frac{z}{\sigma(Z)}; 1\right), \tag{3.48}$$

completing the proof. ♢

Chapter 4
Conclusions and Recommendations for Future Research

In this chapter, the conclusions of the dissertation document are presented, and the possible directions for future research are outlined.

4.1 Conclusions

In this document, we have presented more efficient density estimations for the marginal density of two important semiparametric models. The model discussed in Chapter 2 is considered semiparametric in the sense that the regression function is nonparametric while the error term distribution is parametric. Similarly, the model presented in Chapter 3 is also semiparametric as it assumes a nonparametric structure for the autoregressive conditional variance and a parametric error term distribution.

In Chapter 2, we introduced density and derivative estimators for which it was proved that they could achieve parametric convergence rate, $\sqrt{n}$, and asymptotic normality. We also found that these estimators possess several appealing properties that the classical estimator lacks. Furthermore, we established that in the absence of a nonparametric regression function, in the normal case, the proposed method performs as well as if we have known the model and estimated the density using the maximum likelihood method. Moreover, we also proposed a more powerful density-based specification test for the nonparametric regression function based on the new density estimator. Finally, through extensive numerical studies, we showed that the proposed density estimator, the density derivative estimator, and the specification test significantly outperformed existing ones.

In Chapter 3, we studied the density estimation for the stationary density of nonpara-

metric autoregressive models with conditional heteroscedasticity (ARCH). We demonstrated that the proposed estimator is asymptotically normal and $\sqrt{n}$-consistent, improving this way the widely used classical nonparametric kernel density estimator whose rate of convergence is inferior. In addition, our numerical study revealed that the proposed estimator outperformed this classical estimator. Next, we suggested utilizing our efficient density estimate to get a more accurate estimation for the asymptotic variance of the estimated coefficients in a quantile regression model, where the errors follow our nonparametric ARCH structure. Finally, we conducted a thorough simulation study, which showed that using the new density estimator leads to a more accurate estimation of this asymptotic variance compared to using the classical density estimator.

4.2 Future work directions

As discussed in Chapter 1, density estimation has a fundamental role in Statistics and other fields. Hence, constructing $\sqrt{n}$-consistent density estimates for the marginal density of different semiparametric models is an important task. The following two subsections briefly describe two possible future research directions. Throughout, denote by $\phi\left(y;\mu,\sigma^2\right)$ the normal density of $N\left(\mu,\sigma^2\right)$.

4.2.1 Stationary density estimation in nonparametric time series regression.

Let $\left(X_i,Y_i\right), i\in\mathbb{Z}$, be a real-valued stationary process of interest, $1\leq i\leq n$, from the fully nonparametric nonlinear stochastic regression model of the form

$$Y_i=\mu\left(X_i\right)+\sigma\left(X_i\right)e_i \tag{4.1}$$

where $e_i, i\in\mathbb{Z}$, are independent and identically standard normal distributed (iid) random variables. Here $\mu(\cdot)$ and $\sigma(\cdot)\geq 0$ are measurable and represent drift (or mean regression) and conditional variance (or volatility functions), respectively. Also, in the model (4.1), assume that the observed data is in an increasing time span $n\to\infty$ with a fixed time duration between two consecutive observations. The nonparametric model (4.1) does not impose a specific structures on $\mu(\cdot)$ and $\sigma(\cdot)$, which lets the data speak for themselves and, in some cases, avoid mis-specification.

As mentioned by Zhao and Wu (2008), the model in (4.1) has various special cases. For example, if we let $Y_i = X_{i+1} - X_i$, model (4.1) can be viewed as a discretized version of the stochastic diffusion model

$$dX_i = \mu\left(X_i\right) di + \sigma\left(X_i\right) dW_i, \tag{4.2}$$

where $\{W_i\}$ is a standard Brownian motion. Many well-known financial models are special cases of (4.2); see Fan (2005) and references therein.

Also, the model in (4.1) is a popular model in financial econometrics due to the flexible forms of μ and σ. It allows for nonlinearity and conditional heteroscedasticity. When we let $X_i = Y_{i-1}$, equation (4.1) becomes the conditional heteroskedastic autoregressive nonlinear (CHARN) process $Y_i = \mu\left(Y_{i-1}\right) + \sigma\left(Y_{i-1}\right)\varepsilon_i$ and it can include many parametric time series models as described in Zhao and Wu (2008).

As mentioned in Chapters 2 and 3, the nonparametric kernel density estimator $\tilde{f}_Y(y)$ in (1.1), which is widely used for density estimation, possesses a slow convergence rate $(\sqrt{nb_n})$ that is problematic for small sample sizes. Thus, it is helpful to construct a more efficient density estimation for the marginal density of the nonparametric time series regression model (4.1) to overcome this slow convergence issue of the estimator $\tilde{f}_Y(y)$.

In this possible future work, it is expected to establish similar results as described in Chapters 2 and 3. This is getting an efficient $\sqrt{n}$-consistent density estimate for f_Y by utilizing the specific structure of (4.1). In model (4.1), conditioning on X_i, Y_i has the density function $\phi\left(y; \mu(X_i), \sigma^2\left(X_i\right)\right)$. Thus,

$$f_Y(y) = \mathbb{E}\phi\left(y; \mu(X_i), \sigma^2\left(X_i\right)\right). \tag{4.3}$$

Denote by $(\hat{\mu}(\cdot), \hat{\sigma}^2(\cdot))$ some estimates of $(\mu(\cdot), \sigma^2(\cdot))$. By equation (4.3), the proposed density of $f_Y(y)$ is of the form of

$$\hat{f}_Y(y) = \frac{1}{n}\sum_{i=1}^{n} \phi\left(y; \hat{\mu}(X_i), \hat{\sigma}^2(X_i)\right). \tag{4.4}$$

To implement (4.4), it is necessary to construct estimates $(\hat{\mu}(\cdot), \hat{\sigma}^2(\cdot))$ of $(\mu(\cdot), \sigma^2(\cdot))$. Thus, for $\mu(\cdot)$, use the widely used nonparametric kernel smoothing estimate (Nadaraya-

Watson kernel-weighted)

$$\hat{\mu}_{h_n}(x) = \frac{\sum_{i=1}^{n} Y_i K\left((x - X_i)/h_n\right)}{\sum_{i=1}^{n} K\left((x - X_i)/h_n\right)}, \tag{4.5}$$

for bandwidth $h_n > 0$ and kernel function $K(\cdot)$. Now, for $\hat{\sigma}^2(\cdot)$ in (4.4), also use the Nadaraya-Watson kernel-weighted estimate given by

$$\hat{\sigma}^2_{h_n^*}(x) = \frac{\sum_{i=1}^{n} \left[Y_i - \hat{\mu}_{h_n}(X_i)\right]^2 K\left((x - X_i)/h_n^*\right)}{\sum_{i=1}^{n} K\left((x - X_i)/h_n^*\right)}, \tag{4.6}$$

for proper kernel function $K(\cdot)$ and bandwidth $h_n^* \to 0$. Here h_n^* is another bandwidth, and it can be different from the bandwidth h_n in estimating μ. Finally, for this project, it can be helpful to compare the performance of the proposed density estimator $\hat{f}_Y(y)$ in (4.4) with the classical nonparametric kernel density estimator $\tilde{f}_Y(y)$ in (1.1).

4.2.2 Stationary density estimation in parametric time series models with nonparametric conditional heteroscedasticity.

Consider observations Y_i, $1 \leq i \leq n$, from the heteroscedastic model

$$Y_i = X_i^T \beta + \sigma(U_i)\, e_i \tag{4.7}$$

where $\beta = (\beta_0, \beta_1, \ldots, \beta_{d-1})^T \in \mathbb{R}^{d\times 1}$ is a d-dimensional coefficient vector of the vector of regressors $X_i = (1, X_{1i}, \ldots, X_{(d-1)i}) \in \mathbb{R}^{1\times d}$, $\sigma(\cdot) \geq 0$ is measurable and represent the conditional variance (or volatility functions), $e_i, i \in \mathbb{Z}$, are independent and identically standard normal distributed (iid) random variables, and U_i is an exogenous variable which is independent of vector X_i for all $i = 1, 2, ..., n$. Also, in the model (4.7), assume that the observed data is in an increasing time span $n \to \infty$ with a fixed time duration between two consecutive observations. The model (4.7) does not impose a specific structure on $\sigma(\cdot)$, which lets the data speak for themselves.

Under mild dependence conditions on X_i, it is expected to establish similar results as described in Chapters 2 and 3. This is getting an efficient $\sqrt{n}$-consistent density estimate for f_Y by utilizing the specific structure of (4.7). In model (4.7), conditioning on X_i, Y_i has the density function $\phi\left(y; X_i^T\beta, \sigma^2(U_i)\right)$. Thus,

$$f_Y(y) = \mathbb{E}\phi\left(y; X_i^T\beta, \sigma^2\left(U_i\right)\right). \tag{4.8}$$

Denote by $(\hat{\beta}, \hat{\sigma}^2(\cdot))$ some estimates of $(\beta, \sigma^2(\cdot))$. By equation (4.8), the proposed density of $f_Y(y)$ is of the form of

$$\hat{f}_Y(y) = \frac{1}{n}\sum_{i=1}^{n}\phi\left(y; X_i^T\hat{\beta}, \hat{\sigma}^2(U_i)\right). \tag{4.9}$$

To implement (4.9), it is necessary to construct estimates $(\hat{\beta}, \hat{\sigma}^2(\cdot))$ of $(\beta, \sigma^2(\cdot))$. Under model (4.7), estimate β by the least-squares method:

$$\hat{\beta} = \underset{b\in\mathbb{R}^d}{\operatorname{argmin}}\sum_{i=1}^{n}\left(Y_i - X_i^T b\right)^2. \tag{4.10}$$

For $\sigma^2(\cdot)$, use the widely used nonparametric kernel smoothing estimate (Nadaraya-Watson kernel-weighted)

$$\hat{\sigma}^2_{h_n}(u) = \frac{\sum_{i=1}^{n}\left[Y_i - X_i^T\hat{\beta}\right]^2 K\left((u-U_i)/h_n\right)}{\sum_{i=1}^{n}K\left((u-U_i)/h_n\right)}, \tag{4.11}$$

for proper kernel function $K(\cdot)$ and bandwidth $h_n \to 0$. Similarly to the future work described in Subsection 4.2.1, here it also can be helpful to compare the performance of the proposed density estimator $\hat{f}_Y(y)$ in (4.9), in terms of MSE, with the classical nonparametric kernel density estimator $\tilde{f}_Y(y)$ in (1.1).

Bibliography

Altman, N. S. (1990). Kernel smoothing of data with correlated errors. *Journal of the American Statistical Association*, 85(411):749–759.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858.

Ango-Nze, P. and Doukhan, P. (1998). Functional estimation for time series: Uniform convergence properties. *J. Statist. Plann. Inference*, 68:5–29.

Ango-Nze, P. and Portier, B. (1994). Estimation of the density and of the regression functions of an absolutely regular stationary process. *Pub. I.S. U. P.*, 38:59–87.

Awartani, B. M. A. and Corradi, V. (2005). Predicting the volatility of the S&P 500 stock index via GARCH models: The role of asymmetries. *Int. J. Forecasting*, 21:167–183.

Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, 9:385–426.

Aït-Sahalia, Y., Fan, F., and Peng, H. (2009). Nonparametric transition-based tests for jump-diffusions. *Journal of the American Statistical Association*, 104:1102–1116.

Bickel, P. and Ritov, Y. (1988). Estimating integrated squares density derivatives: sharp best order of convergence estimates. *Sankhya, Series A*, 50:381–393.

Birgé, L. and Massart, P. (1995). Estimation of integral functionals of a density. *Annals of Statistics*, 23:11–29.

Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994). Handbook of econometrics. In *Chapter 49 Arch models*, volume 4, pages 2959–3038. Elsevier.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*. 2nd ed. Lecture Notes in Statistics 110, New York:Springer.

Caporale, G. M. and Gil-Alana, L. A. (2004). Long range dependence in daily stock returns. *Appl. Financ. Econ*, 14:375–383.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. 2nd Edition. Belmont, CA, Duxbury.

Chan, K., Kot, H. W., and Tang, G. Y. (2013). A comprehensive long-term analysis of s&p 500 index additions and deletions. *Journal of Banking and Finance*, 37(12):4920–4930.

Dedecker, J., Doukhan, P., Lang, G., León, J. R., Louhichi, S., and Prieur, C. (2007). *Weak dependence: With examples and applications*. Lecture Notes in Statistics 190.

Delaigle, A., Meister, A., and Rombouts, J. (2016). Root-t consistent density estimation in garch models. *Journal of Econometrics*, 192(1):55–63.

Ding, Z., Granger, C. W. J., and Engle, R. F. (1993). A long memory property of stock market returns and a new model. *J. Empirical Finance*, 1:83–106.

Doukhan, P. and Louhichi, S. (2001). Functional estimation of a density under a new weak dependence condition. *Scand. J. Stat*, 28:325–341.

Engle, R. (2001). Garch 101: The use of arch/garch models in applied econometrics. *Journal of Economic Perspectives*, 15(4):157–168.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, 87:998–1004.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Annals of Statistics*, 21:196–216.

Fan, J. (2005). A selective overview of nonparametric methods in financial econometrics. *Statist. Sci.*, 20:317–337.

Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *Annals of Statistics*, 20:2008–2036.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.

Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, 85(3):645–660.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer-Verlag, New York.

Frees, E. W. (1994). Estimating densities of functions of observations. *Journal of the American Statistical Association*, 89:517–525.

Gao, J. and King, M. (2004). Adaptive testing in continuous-time diffusion models. *Econometric Theory*, 20:844–882.

Giné, E. and Mason, D. M. (2007). On local U-statistic processes and the estimation of densities of functions of several sample variables. *Annals of Statistics*, 35:1105–1145.

Gutenbrunner, G. and Jurečková, J. (1992). Regression rank scores and regression quantiles. *Annals of Statistics*, 20:305–330.

Hall, P. and Marron, J. (1987). Estimation of integrated squared density derivatives. *Statistics & Probability Letters*, 6:109–115.

Hansen, B. E. (1992). Consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica*, 60(4):967–972.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(3):726–748.

Herrmann, E., Gasser, T., and Kneip, A. (1992). Choice of bandwidth for kernel regression when residuals are correlated. *Biometrika*, 79(4):783–795.

Hong, Y. and Li, H. (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Review of Financial Studies*, 18:37–84.

IMF (2004). *Consumer price index manual: Theory and practice - International Monetary Found*. https://www.ilo.org/public/english/bureau/stat/download/cpi/cpi-manual-en.pdf.

Jennrich, R. (1969). Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics*, 40:633–643.

Jones, M. C. (1993). Simple boundary correction for kernel estimation. *Statistics and Computing*, 3(3):135–146.

Kerkyacharian, G. and Picard, D. (1996). Estimating nonquadratic functionals of a density using Haar wavelets. *Annals of Statistics*, 24:485–507.

Kim, K. and Wu, W. B. (2007). *Density estimation for nonlinear time series*. Manuscript.

Kim, K. H., Zhang, T., and Wu, W. B. (2015). Parametric specification test for nonlinear autoregressive models. *Econometric Theory*, 31(5):1078–1101.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press, Cambridge.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Koenker, R., Chernozhukov, V., He, X., and Peng, L. E. (2018). *Handbook of Quantile Regression, 1st ed.* Chapman and Hall/CRC.

Koenker, R. and Zhao, Q. (1994). L-estimatton for linear heteroscedastic models. *Journal of Nonparametric Statistics*, 3:223–235.

Koenker, R. and Zhao, Q. (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory*, 12:793–813.

Koenker, R. W. and D'Orey, V. (1987). Computing regression quantiles. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 36(3):383–393.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. *Journal of Econometrics*, 54:159–178.

Li, Q. and Racine, J. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Li, S. and Tu, Y. (2016). Root n consistent density estimation in semiparametric regression models. *Computational Statistics & Data Analysis*, 104:91–109.

Li, X. and Zhao, Z. (2019). A Time Varying Approach to the Stock Return–Inflation Puzzle. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 68(5):1509–1528.

Liao, Y. and Stachurski, J. (2015). Simulation-based density estimation for time series using covariate data. *Journal of Business & Economic Statistics*, 33(4):595–606.

Müller, H.-G. (1993). On the boundary kernel method for non-parametric curve estimation near endpoints. *Scandinavian Journal of Statistics*, 20(4):313–328.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Perron, P. (1988). Trends and random walks in macroeconomic time series. *Journal of Economic Dynamics and Control*, 12:297–332.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, 7(2):186–199.

Pritsker, M. (1998). Nonparametric density estimation and tests of continuous time interest rate models. *Review of Financial Studies*, 11:449–487.

Racine, J. (2008). Nonparametric econometrics: A primer. *Foundations and Trends(R) in Econometrics*, 3(1):1–88.

Ritov, Y. and Bickel, P. (1990). Achieving information bounds in non and semiparametric models. *Annals of Statistics*, 18:925–938.

Robinson, P. M. (1983). Nonparametric estimators for time series. *J. Time Series Anal*, 4:185-207.

Roussas, G. G. (2000). Asymptotic normality of the kernel estimate of a probability density function under association. *Statist. Probab. Lett*, 50:1–12.

Rudin, W. (1976). *Principles of Mathematical Analysis*. 3rd Edition. McGraw-Hill.

Ruppert, D., Sheather, S. J., and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, 90:1257–1270.

Saavedra, A. and Cao, R. (1999). Rate of convergence of a convolution-type estimator of the marginal density of an MA(1) process. *Stochastic Processes and their Applications*, 80:129–155.

Saavedra, A. and Cao, R. (2000). On the estimation of the marginal density of a moving average process. *Canadian Journal of Statistics*, 28:799–815.

Schick, A. and Wefelmeyer, W. (2004). Functional convergence and optimality of plug-in estimators for stationary densities of moving average processes. *Bernoulli*, 10:889–917.

Schick, A. and Wefelmeyer, W. (2007). Uniformly root-n consistent density estimators for weakly dependent invertible linear processes. *Annals of Statistics*, 35:815–843.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

Teräsvirta, T. and Zhao, Z. (2011). Stylized facts of return series, robust estimates and three popular models of volatility. *Applied Financial Economics*, 21(1-2):67–94.

Truquet, L. (2019). Root-n consistent estimation of the marginal density in semiparametric autoregressive time series models. *Bernoulli*, 25(3):2107 – 2136.

Wand, M. P. and Jones, M. C. (1994). *Kernel Smoothing*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability (60), Boca Raton, FL, U. S.

White, H. (1984). *Asymptotic Theory for Econometricians*. Academic Press, New York.

Wu, W. B. and Zhao, Z. (2007). Inference of Trends in Time Series. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(3):391–410.

Zhao, Q. (2001). Asymptotically efficient median regression in the presence of heteroskedasticity of unknown form. *Econometric Theory*, 17(4):765–784.

Zhao, Z. (2008). Parametric and nonparametric models and methods in financial econometrics. *Statistics Surveys*, 2:1–42.

Zhao, Z. (2010). Density estimation for nonlinear parametric models with conditional heteroscedasticity. *Journal of Econometrics*, 155:71–82.

Zhao, Z. (2011). Nonparametric model validations for hidden Markov models with applications in financial econometrics. *Journal of Econometrics*, 162.:225–239.

Zhao, Z. and Wu, W. B. (2008). Confidence bands in nonparametric time series regression. *The Annals of Statistics*, 36(4):1854 – 1878.

Vita

Manuel Dario Hernandez Bejarano

Education:

May 2024	Ph.D. in Statistics, **The Pennsylvania State University**.
Mar 2017	M.Sc. in Statistics, **The National University of Colombia**.
Oct 2009	B.Sc. in Statistics, **The National University of Colombia**.
Apr 2007	B.Sc. in Physics, **The Universidad Distrital Francisco Jose de Caldas**, Colombia.

Professional Experience

- Aug 2019-May 2024: Graduate Teaching Assistant, *The Pennsylvania State University*, Department of statistics.
- Jul 2018-Jul 2019: Instructor, *The Externado de Colombia University*, Faculty of Finance and Business. Master in Finance.
- Jun 2013-Jul 2019: Specialized Professional, *Central Bank of Colombia*, Technical and Economic Information Department.
- Jul 2009-May 2013: Professional in Statistics, *Central Bank of Colombia*, Technical and Economic Information Department.
- Jun 2008-Jun 2009: Statistical Analyst, *Central Bank of Colombia*, Technical and Economic Information Department.
- Sep 2007-Jun 2008: Research assistant, *The National University of Colombia*, Department of statistics.

Skills

- Proficient with: time series analysis, analysis of stochastic process, forecast, non-parametric statistics, machine learning, Bayesian econometrics, multivariate statistics, and generalized linear models.
- Proficient with: R, SAS, Stan, Python, Matlab, Minitab, GAUSS, RATS, LaTeX.

.