Forecasting the Confirmed COVID-19 Cases Using Modal Regression

XIN JING
School of Economics, Yonsei University, Seodaemun-gu, Seoul 03722, South Korea
Email: superheunheun@gmail.com

JIN SEO CHO
School of Economics, Yonsei University, Seodaemun-gu, Seoul 03722, South Korea
Email: jinseocho@yonsei.ac.kr

This version: December 2023

Abstract

This study utilizes modal regression to forecast the cumulative confirmed COVID-19 cases in Canada, Japan, South Korea, and the United States. The objective is to improve the accuracy of the forecasts compared to standard mean and median regressions. To evaluate the performance of the forecasts, we conduct simulations and introduce a metric called the coverage quantile function (CQF), which is optimized using modal regression. By applying modal regression to popular time-series models for COVID-19 data, we provide empirical evidence that the forecasts generated by the modal regression outperform those produced by the mean and median regressions in terms of the CQF. This finding addresses the limitations of the mean and median regression forecasts.

Key Words: Forecasting COVID-19 cases; Modal regression; Conditional mode; MEM algorithm; Density estimation.

JEL Classifications: C22, C53, I18.

Acknowledgements: The editor-in-chief, Derek Bunn, and an anonymous referee provided very helpful comments for which we are most grateful. The authors are grateful to Jihye Jung, Jaeseung Lee, and Moo Hyun Yang for their constructive comments. Jing and Cho acknowledge research support from the Yonsei University Research Grant of 2023.

1 Introduction

Since the first confirmed identification of corona virus disease (COVID-19) in Wuhan, China, in December 2019, it has emerged as a global concern. On March 11, 2020, the World Health Organization (WHO) declared it a global pandemic. By April 2022, there were over 500 million confirmed cases and 6 million deaths worldwide. Given the seriousness of COVID-19, numerous countries swiftly implemented various measures, including short-term actions, such as lockdowns, and long-term strategies, such as social distancing and vaccine development.

The economic impact of COVID-19 has been severe, leading governments worldwide to respond with a significant increase in budget expenditures. As depicted in Table 1, governments have employed various measures to combat COVID-19, resulting in substantial budget allocations. For instance, the U.S. and Canadian economies witnessed a 2.14% and 1.99% increase in health expenditure between 2019 and 2020, respectively. Consequently, both the U.S. and Canada were able to provide the public with vaccines starting in December 2020, while the Japanese and South Korean governments followed suit in February 2021. By December 2022, the vaccination rates per 100 people reached 247.84, 285.44, 250.19, and 197.37 for Canada, Japan, South Korea, and the U.S., respectively. These vaccination efforts led to reductions in the fatality rates to 1.1%, 0.2%, 0.1%, and 1.1% for the respective countries.

<Insert Table 1 around here>

Governmental health policies are formulated to anticipate the future course of a pandemic. The effectiveness of government responses to a pandemic relies heavily on the accuracy of these forecasts. Accurate pandemic forecasting allows for the early identification and mitigation of potential health risks, thereby preventing the spread and outbreak of the disease. Additionally, forecasting plays a crucial role in enabling healthcare systems and governments to efficiently plan and allocate resources, such as medical equipment, personnel, and drugs. This ensures optimal control and mitigation of the disease's impact. For instance, if the forecasted number of confirmed COVID-19 cases is lower than the actual number, governments may face the risk of increasing fatality rates owing to limited vaccine availability from suppliers. Therefore, accurate forecasting is vital for governments to make informed decisions on the procurement and distribution of vaccines and other essential resources.

Therefore, the primary objective of our study is to introduce an alternative methodology for forecasting COVID-19 cases that complements the forecasts generated by the standard mean and median regression methods. To achieve this goal, we leverage the benefits offered by modal regression compared to mean and median regressions. Our study focuses on forecasting the cumulative confirmed COVID-19 cases for the

four countries mentioned above using the modal regression approach. By utilizing this methodology, we aim to enhance the accuracy and reliability of COVID-19 forecasts, providing a valuable addition to the existing forecasting methods based on mean and median regressions.

A considerable body of literature has been dedicated to empirically analyzing the COVID-19 trend by applying various classical forecasting methods. Studies such as Boccaletti, Ditto, Mindlin, and Atangana (2020); Almeshal, Almazrouee, Alenizi, and Alhajeri (2020); Vespignani, Tian, Dye, Lloyd-Smith, Eggo, Shrestha, Scarpino, Gutierrez, Kraemer, Wu et al. (2020); Chan, Chu, Zhang, and Nadarajah (2021); Gning, Ndour, and Tchuenche (2022) focus on estimating the strain on medical services, understanding epidemiological patterns, and providing policymakers with comprehensive information to formulate effective policies. Furthermore, Musulin, Baressi Šegota, Štifanić, Lorencin, Anđelić, Šušteršič, Blagojević, Filipović, Ćabov, and Markova-Car (2021) review the application of standard regression methods in various AI-based COVID-19 applications. However, conditional mean- or median-based forecasts can be significantly influenced by outliers or heavy-tailed noise in the data (see Chen, Genovese, Tibshirani, and Wasserman, 2016; Zhou and Huang, 2016; Xiang and Yao, 2022, for example). These limitations highlight the need for alternative forecasting approaches to mitigate the impact of such data irregularities.

Modal regression is a valuable alternative to standard regression methods for forecasting random processes that contain outliers and/or exhibit heavy-tailed noise distributions. The literature on modal regression has demonstrated its advantages over other forecasting techniques. Sasaki, Sakai, and Kanamori (2020) recognize that estimating the conditional mode is more robust than estimating the conditional mean or median, particularly when dealing with wide-ranging noise. Xiang and Yao (2022) provide an intuitive location estimator for skewed data, highlighting the superiority of modal regression in such cases. Furthermore, Yao and Li (2014) introduce modal linear regression, exploring its application to high-dimensional data and analyzing its asymptotic properties without assuming a symmetric error density function. Yu, Zhu, Shi, and Ai (2020) propose a robust estimation procedure for partial functional linear regression using modal regression, specifically designed to handle outliers and heavy-tailed error distributions. Xiang and Yao (2022) also propose a novel nonparametric statistical learning tool based on modal regression, serving as a complementary approach to standard mean and median regressions. Overall, the literature highlights the advantages of modal regression in addressing the challenges posed by outliers, heavy-tailed noise, and skewed data, thereby providing a robust and flexible forecasting methodology.

Despite the recent advancements in modal regression, its application to COVID-19 data remains limited. The COVID-19 pandemic has exhibited exponential growth, a wide range of government responses, and the emergence of unexpected virus variants. These factors introduce unexpected noise and outliers,

making COVID-19 data an excellent opportunity to explore the capabilities of modal regression. Although Ullah, Wang, and Yao (2022) have conducted preliminary investigations by applying modal regression to examine the interrelationship between COVID-19 cases and deaths in the U.S., there is still scope for further exploration using modal regression. In this study, we utilize modal regression to forecast COVID-19 data using specific time-series models. We then compare the results with those obtained using mean and median regressions, as highlighted in previous studies (e.g., Musulin et al., 2021).

This study contributes to the existing literature in two significant ways. First, we introduce the coverage quantile function (CQF) as a metric to evaluate the performance of modal regression. While the root-mean-squared error and mean absolute error are commonly used objectives for optimizing mean and median regressions, respectively, we utilize CQF as the objective for the modal regression. This approach provides a clear understanding of the role and effectiveness of modal regression in forecasting. Second, we specify an autoregressive model to capture the serial correlation in the COVID-19 data for the four countries listed in Table 1. We then apply modal regression to generate forecasts and compare them with the forecasts obtained using mean and median regressions. Our analysis reveals that the modal regression outperforms the mean and median regressions in forecasting outliers. This finding underscores the superior performance of modal regression in handling the unique characteristics and challenges of COVID-19 data. Overall, this study's contributions lie in the introduction of CQF as a novel evaluation metric for modal regression and the empirical demonstration of its superior forecasting capabilities compared to mean and median regressions, particularly when dealing with outliers.

The methodology employed in this study involves a simulation approach. The implementation of the modal regression method relies on estimating the conditional density function, which is highly sensitive to factors such as the selection of bandwidth or the shape of the density function, as pointed out by Ullah et al. (2022). Consequently, the valuable theoretical results regarding modal regression are often challenging to validate using empirical data owing to the presence of irregular data patterns. To address this challenge, we adopt an extensive simulation approach that allows us to examine the characteristics of modal regression using finite samples. We conduct Monte Carlo simulations using both cross-sectional and time-series data to evaluate the performance of mean, median, and modal regressions. Additionally, we compare the performance of different bandwidths utilized in the modal regression estimation. Furthermore, we apply the modal regression method to forecast the cumulative confirmed COVID-19 cases in the four countries mentioned earlier. Through this empirical application, we demonstrate that the modal regression-based forecast achieves a superior CQF compared with other forecasting methods. By combining simulation studies and empirical analysis, we assess the performance of modal regression under various scenarios, investigate the

impact of different bandwidth choices, and highlight the advantages of modal regression in forecasting COVID-19 cases.

The remainder of this paper is organized as follows. In Section 2, we present a comprehensive review of the relevant literature related to the subject of this study, highlighting the motivation behind our research. Section 3 focuses on formalizing the modal regression problem. We propose the CQF metric and provide an overview of the existing modal regression methods. The simulation results are presented in Section 4, in which various simulations are conducted to evaluate the performance of mean, median, and modal regressions. Section 5 discusses the empirical analysis applied to COVID-19 data. We apply the modal regression method to forecast the cumulative confirmed COVID-19 cases for the four countries mentioned earlier and compare the results with those of other forecasting approaches. Finally, in Section 6, we conclude the study by summarizing the key findings and discussing the implications and potential future directions of research in this field.

2 Literature Review and Motivation

Numerous studies have focused on analyzing and predicting the trends of COVID-19. This section provides a brief overview of some of the methodologies employed in these studies.

One common approach is to develop empirical prediction models using machine learning methods. For example, Car, Baressi Šegota, Anđelić, Lorencin, and Mrzljak (2020) train a multilayer perceptron (MLP) artificial neural network to create a global model for forecasting the maximum number of patients across various locations over time. Similarly, Mollalo, Rivera, and Vahedi (2020) utilize MLP to forecast the cumulative COVID-19 incidence rates, specifically in the U.S. Chakraborty and Ghosh (2020) propose a hybrid approach that combines integrated autoregressive moving average models with wavelet-based forecasting models to predict the number of daily confirmed cases in the short term. Other prediction models for confirmed cases include the gradient boosting regression model, the generalized waring regression model, and various other machine learning approaches (see also Gumaei, Al-Rakhami, Al Rahhal, Albogamy, Al Maghayreh, and AlSalman, 2021; Gning et al., 2022, for more examples using machine learning methods). These studies highlight the versatility of machine learning methods in capturing the complex dynamics of COVID-19 data and providing accurate predictions. By leveraging various machine learning algorithms, researchers have made significant strides in understanding and forecasting the spread of the virus.

In addition to machine learning methods, evolutionary computing algorithms have been employed to develop epidemiological models that capture biological evolution through processes such as reproduction,

mutation, recombination, and selection. For instance, Salgotra, Gandomi, and Gandomi (2020a) utilize gene expression programming (GEP) based on evolutionary data analysis to specify a model for the potential impact of COVID-19 on the 15 most affected countries. Similarly, in another study by Salgotra, Gandomi, and Gandomi (2020b), a robust and reliable variant of the GEP method is developed to model the confirmed cases and deaths caused by COVID-19 in India. Other examples include the work by Yousefpour, Jahanshahi, and Bekiros (2020), who propose an effective and efficient multi-objective genetic algorithm for designing government strategies to address the disease. Zivkovic, Bacanin, Djordjevic, Antonijevic, Strumberger, Rashid et al. (2021) employ a hybrid model combining an adaptive neuro-fuzzy inference system and an enhanced genetic algorithm to predict the number of confirmed cases in China. These studies demonstrate the application of evolutionary computing algorithms in modeling the dynamics of the COVID-19 pandemic, providing valuable insights and predictions. By incorporating evolutionary principles, these approaches offer unique perspectives and the potential to optimize strategies for mitigating the impact of the virus.

Several studies have focused on analyzing the COVID-19 trend from an economic perspective and assessing its impact on the economy. They provide valuable insights into various economic implications of the pandemic. For example, Almeshal et al. (2020) investigate the effectiveness of non-pharmaceutical intervention measures in forecasting the size of the COVID-19 pandemic in Kuwait. They employ deterministic and stochastic modeling approaches to estimate the scale of confirmed COVID-19 cases and identify the ending phase of the pandemic. Their findings highlight the efficacy of non-pharmaceutical interventions, particularly when infection rates and personal contact patterns change over time. Other studies examine specific economic impacts of COVID-19. Ajide, Ibrahim, and Alimi (2020) analyze the impact of the lockdown policy implementation on confirmed COVID-19 cases in Nigeria. Azimli (2020) investigate the impact of COVID-19 on the degree and dependence structure of risky asset returns in the U.S. Béland, Brodeur, and Wright (2023), Gupta, Montenovo, Nguyen, Lozano-Rojas, Schmutte, Simon, Weinberg, and Wing (2023), and Rojas, Jiang, Montenovo, Simon, Weinberg, and Wing (2020) examine the effects of COVID-19 on the labor market. Furthermore, Lu, Nie, and Qian (2021), Hamermesh (2020), Béland, Brodeur, Mikola, and Wright (2022), and Tubadji, Boy, and Webber (2020) explore the impact of COVID-19 on mental health and well-being, while Olmstead and Tertilt (2020) delves into a detailed examination of the impact of COVID-19 on gender inequality. Studies by Andrée (2020), He, Pan, and Tanaka (2020), Brodeur, Cook, and Wright (2021), and Almond, Du, and Zhang (2020) investigate the environmental effects of COVID-19. These studies provide valuable insights into the multifaceted economic consequences of COVID-19 and the corresponding government responses. They shed light on the impact on sectors such as labor markets, mental health, gender equality, and the environment. For a comprehensive review of the economic consequences of

COVID-19 and government responses, Brodeur et al. (2021) provide a recent survey.

Despite the extensive research conducted on COVID-19 and its analysis, there are variations in the generated forecasts, and they may not be directly applicable to forecasting economic activities. Existing literature often predicts the COVID-19 trend using mean and median regressions. For example, Rojas et al. (2020) and Hamermesh (2020) employ mean regression to forecast the impact of COVID-19, whereas Lu et al. (2021) delve deeper into mean-based forecasts using a median regression. Additionally, Béland et al. (2022), Gupta et al. (2023), Tubadji et al. (2020), and He et al. (2020) apply the difference-in-differences approach to evaluate the COVID-19 policies. However, forecasting the peak of the pandemic could be more relevant when it comes to forecasting the economic environment affected by the pandemic. Economic activities before and after the peak are likely to differ significantly, making it important to accurately forecast the peak. Mean and median regressions may not be suitable for this purpose because they assume the central tendency of the conditional distribution through mean and median estimations, respectively. Moreover, the mean regression is most efficient when the conditional distribution is Gaussian or sub-Gaussian, whereas the median regression becomes a robust estimator when the distribution is light-tailed. Unless the distribution of confirmed cases is unimodal and symmetric, the mean and median regressions struggle to capture the most likely value of the conditional distribution–a challenging characteristic often observed in real-world data. Real-world data are more likely to exhibit multimodal, skewed, or fat-tailed distributions. Studies by Krief (2017) and Ullah, Wang, and Yao (2021) demonstrate that mean and median regressions lose their robustness and/or efficiency when time-series datasets contain multiple outliers and/or skewed distributions.

This aspect serves as the motivation to forecast the peak by estimating the mode of the conditional distribution. To achieve this, we use modal regression, which is specifically designed to estimate the mode of a conditional distribution. In addition to estimating the same quantities as in the mean and median regressions, under the assumption of a unimodal and symmetric conditional distribution, modal regression offers several additional properties. First, modal regression is more robust to outliers than the mean and median estimators because it utilizes the mode as a representation of the central tendency of the conditional distribution. Second, modal regression yields narrower forecasting intervals than other estimations. This is because the interval around the conditional mode contains more observations than that around the conditional mean and/or median for the same interval size. Finally, for data drawn from a multimodal conditional distribution, modal regression captures a central tendency that is different from the mean and/or median. This enables the exploration of different aspects of the conditional distribution. In this study, we apply modal regression to forecast confirmed COVID-19 cases in Canada, Japan, South Korea, and the U.S. Subsequently, we compare these forecasts with those obtained through mean and median regressions to assess the performance

and advantages of the modal regression approach.

3 Method of Modal Regression

In this section, we discuss the methodology of modal regression and the models used for forecasting. Additionally, we provide a comprehensive explanation of the criteria used to evaluate the forecasts generated by the modal regression.

3.1 Modal Regression

First, we discuss the limitations of the mean regression for forecasting and compare its characteristics with those of the modal regression. To begin, let $\mathcal{X} \subset \mathbb{R}^d$ and $\mathcal{Y} \subset \mathbb{R}$ represent the spaces of regressors and the dependent variable, respectively. We consider a dataset $\mathcal{S} = \{(x_t, Y_t) \in \mathcal{X} \times \mathcal{Y} : t = 1, 2, \ldots, n\}$ consisting of independent and identically distributed random samples.[1] Given this setup, we define the conditional mode function $f(x)$ as the mode of the conditional probability density function $p_{Y|X}(\cdot)$, which represents the density of $Y \in \mathcal{Y}$ conditioned on $X \in \mathcal{X}$. Mathematically, we have

$$f(x) := \text{mode}(Y|X = x) := \arg\max_y p_{Y|X}(y|X = x).$$

Additionally, we introduce the variable U, defined as the difference between Y and $f(X)$, that is, $U := Y - f(X)$. Under the assumption that $p_{U|X}(\cdot|X = x)$ is continuous and bounded for any $X = x \in \mathcal{X}$, the conditional mode of U given $X = x$ is denoted as $\text{mode}(U|X = x)$, which satisfies $\arg\max_y p_{U|X}(y|X = x) = 0$. We assume that the conditional density function $p_{Y|X}(\cdot|X = x)$ has a unique maximum with a probability of 1, ensuring that $f(x)$ is well-defined for any $x \in \mathcal{X}$. This assumption guarantees the existence and uniqueness of the global mode of $f(\cdot)$.[2] Furthermore, maximizing the conditional density $p_{Y|X}(\cdot|X = x)$ is equivalent to maximizing the joint density $p_{X,Y}(X = x, Y = \cdot)$ because $p(y|x) = p(x, y)/p(x)$ for a fixed $x \in \mathcal{X}$. Therefore, the conditional mode can also be expressed as $\arg\max_y p(y, x)$, indicating that the estimation of the conditional mode is related to the estimation of the density function of the random variables.

We explore this aspect by representing the conditional mode using a parametric model. Specifically,

[1] It is important to note that COVID-19 data are non-stationary and typically exhibit serial correlation in the error terms, which violates the assumption of independence and identical distribution. We will address this issue later by specifying a model for serial correlation.

[2] This type of conditional mode function is known as the unimodal regression function (e.g., Chen, 2018), while Chen et al. (2016) relaxes the uniqueness condition in a nonparametric model regression context. In this study, we focus on the unimodal case.

we let $\mathcal{M} := \{m(X, \theta) : \theta \in \Theta\}$ be a parametric model for the conditional mode such that for a unique $\theta_* \in \Theta$, mode$(Y|X) = m(X, \theta_*)$, where Θ is a compact parameter space in $\mathbb{R}^p$, and for each $\theta \in \Theta$, $m(\cdot, \theta)$ is a measurable function. Given this parametric model assumption, we estimate θ_* by maximizing the kernel-based objective function:

$$Q_{n,h}(\cdot) := \frac{1}{n}\sum_{t=1}^{n}\phi_h\left(Y_t - m(X_t, \cdot)\right),$$

where for each $u \in \mathbb{R}$, $\phi_h(u) = h^{-1}\phi(u/h)$ such that $\phi(\cdot)$ is a kernel density function symmetric around zero and $\int \phi(u)du = 1$, and h is the bandwidth. The commonly used kernel functions include Gaussian, Epanechnikov, uniform, and triangular functions. For the remainder of this paper, we assume that $\phi(\cdot)$ is the standard normal density function for simplicity. By maximizing the objective function $Q_{n,h}(\cdot)$, we can estimate the value of θ_* corresponding to the conditional mode. This approach allows us to represent the conditional mode using a parametric model and estimate the associated parameters.

The maximization of $Q_{n,h}(\cdot)$ requires a numerical optimization procedure because its maximum has no closed-form expression. Various numerical optimization algorithms can be employed to maximize $Q_{n,h}(\cdot)$ such as the modal expectation and maximization (MEM), Newton-type, and mean-shift algorithms (see Yao and Li, 2014; Khardani and Yao, 2017; Chen et al., 2016, respectively). Among these optimization methods, the MEM algorithm has exhibited robust performance, as demonstrated in the simulations outlined in Section 4. Therefore, for our empirical applications, we utilize the MEM algorithm as the chosen numerical optimization procedure to maximize $Q_{n,h}(\cdot)$ and estimate the parameters associated with the conditional mode.

The MEM algorithm proposed by Li, Ray, and Lindsay (2007) extends the EM algorithm proposed by Dempster, Laird, and Rubin (1977) to the modal regression context. While the EM algorithm assumes the presence of latent variables in the likelihood function, the MEM algorithm considers their presence in the density function and estimates the unknown parameters using the E- and M-steps. Specifically, the E-step involves computing the weight of each observation. Given an initial parameter $\theta^{(0)}$, for each observation $(Y_t, X_t')'$, we calculate

$$\pi(t|\theta^{(0)}) := \frac{\phi_h(Y_t - m(X_t; \theta^{(0)}))}{\sum_{i=1}^{n}\phi_h(Y_t - m(X_t; \theta^{(0)}))}.$$

Next, the M-step maximizes the objective function:

$$\theta^{(1)} := \arg\max_{\theta\in\Theta}\sum_{t=1}^{n}\left\{\pi(t|\theta^{(0)})\log\phi_h\left(Y_t - m(X_t; \theta)\right)\right\}.$$

Using the updated parameter $\theta^{(1)}$, we compute $\pi(t|\theta^{(1)})$ for each t and repeat the maximization process, replacing $\theta^{(0)}$ in the objective function. We iterate the E- and M-steps until the maximizing parameters converge. Denoting the converged parameter as $\widehat{\theta}_n$, it maximizes the objective function $Q_{n,h}(\cdot)$ since each iteration progressively maximizes $Q_{n,h}(\cdot)$. In the Appendix, we prove that for any positive integer k, $Q_{n,h}(\theta^{(k+1)}) - Q_{n,h}(\theta^{(k)}) \geq 0$. Therefore, as k tends to infinity, the maximum value of $Q_{n,h}(\cdot)$ is reached. This proof remains valid even when $m(X_t;\cdot)$ is nonlinear, thereby generalizing the proof in Yao and Li (2014), which assumes a linear model of the conditional mode.

The MEM algorithm relies on estimating the objective function using the kernel density function estimation, which is influenced by the choice of bandwidth h. However, as highlighted by Ullah et al. (2022) and confirmed by our Monte Carlo simulations in Section 4, the convergence of $\widehat{\theta}_n$ to the unknown true parameter depends critically on the bandwidth selection. Among various bandwidth selection methods, the bandwidth suggested by Sheather and Jones (1991), referred to as SJ, generally produces robust estimation results along with other bandwidths such as those selected by Scott's (1979) and Silverman's (1986) rule of thumb. The SJ's bandwidth denoted as h_n^{SJ} is given as follows. Suppose $z_1, z_2, \ldots, z_n$ represents the sample points of a random variable Z. Then,

$$h_n^{SJ} := \left(\frac{\int k^2(u)du}{n\widehat{\sigma}_n^4 \int \left[\widehat{f}_n''(u)\right]^2 du} \right)^{\frac{1}{5}},$$

where $\widehat{\sigma}_n$ is the standard deviation estimated using the sample points, $k(\cdot)$ is a kernel function used to weigh the sample points, and $\widehat{f}_n''(\cdot) := \frac{1}{nh_0^3} \sum_{t=1}^{n} L'' \left(\frac{(\cdot)-z_t}{h_0} \right)$ estimates $f''(\cdot)$, the second derivative of the density function $f(\cdot)$ of Z. Here, h_0 is the bandwidth, and $L(\cdot)$ is the kernel function used to estimate $f(\cdot)$. SJ suggests the use of a simple rule of thumb for h_0. In addition to SJ's bandwidth, two other commonly used Scott's and Silverman's bandwidths are as follows:

$$h_n^{SC} := 1.06\widehat{\sigma}_n n^{-1/5} \quad \text{and} \quad h_n^{SV} := 0.9 \min[\widehat{\sigma}_n, IQR_n/1.34] n^{-1/5},$$

respectively, where IQR_n represents the interquartile range of the sample points, that is, the distance between the second and third quartiles.

When applying the MEM algorithm, we combine it with least squares estimation to first estimate the density function. Here is the procedure we follow when using SJ's bandwidth as an example: first, we begin by estimating the standard deviation, denoted as $\widehat{\sigma}_n^{(1)}$, using the residuals obtained from the least squares estimation. Next, we apply the MEM algorithm to the density function estimated with SJ's bandwidth

using $\widehat{\sigma}_n^{(1)}$, and obtain the first-step MEM estimator, denoted as $\theta^{(1)}$. Using $\theta^{(1)}$, we compute different residuals to estimate the standard deviation, denoted as $\widehat{\sigma}_n^{(2)}$, and estimate the density function using $\widehat{\sigma}_n^{(2)}$. Subsequently, we maximize this new density function and obtain the second-step MEM estimator, denoted as $\theta^{(2)}$. We continue this iterative process, estimating $\widehat{\sigma}_n^{(k)}$ and obtaining the k-th step MEM estimator $\theta^{(k)}$ until convergence is reached, and obtain $\widehat{\theta}_n$. We propose estimating $\widehat{\sigma}_n^{(1)}$ using least-squares estimation because it is not straightforward to estimate $\widehat{\sigma}_n$ directly using the MEM algorithm when the data are nonstationary. In such cases, it is useful to first estimate the conditional mean using a unit-root process, as demonstrated by the simulation in Section 4.2.

The main objective of modal regression differs from that of mean and median regressions. Although the mean squared error (MSE) and mean absolute error (MAE) are typically used as target metrics for optimizing mean and median regressions, respectively, these metrics do not align with the objective of the conditional mode function (see, e.g., Buhai, 2005; Porter, 2015). Therefore, for modal regression, a different target metric is required. Given that the conditional mode reflects the density in the vicinity of $f(x)$ directly, we can define a natural objective metric based on the number of observations around the estimator. In this context, the CQF can serve as an appropriate objective metric for modal regression. Specifically, for a given $\tau \in (0,1)$, if κ is the quantity satisfying the equality

$$\mathbb{E}[\mathbb{I}(|Y - g(X)| \leq \kappa)] = \tau,$$

where $\mathbb{I}(\cdot)$ is the indicator function and $g(X)$ is a quantity defined by X, we can characterize the behavior of the conditional mode using κ. Intuitively, as κ decreases, $g(X)$ should approach the conditional mode $f(X)$, indicating that for a fixed τ, we can estimate κ to measure the extent to which the conditional density of $Y|X$ is concentrated around $g(X)$. Therefore, we define the sample analog for the CQF as:

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(|Y_t - \widehat{g}_n(X_t)| \leq \kappa) = \tau,$$

where $\widehat{g}_n(X_t)$ is a generic estimator. This equation allows us to determine κ that satisfies the equality, providing a measure of how well the estimator captures the concentration of the conditional density around $g(X)$. For the empirical applications discussed in Section 5, we set $\tau = 0.50$, meaning that the interval around $g(X_t)$ with a distance of 2κ covers half the observations in the dataset. When $\widehat{g}_n(X_t)$ estimates the conditional mode, we can expect the interval characterized by κ to be narrower than that those characterized

by the conditional mean or median.[3]

3.2 Extension of the Modal Regression

In this section, we extend the application of modal regression to time-series data. This is necessary because the data assumption made in the previous section does not account for the presence of a serial correlation in the cumulative confirmed COVID-19 cases. To address this issue, we introduce time-series models that explicitly capture the serial correlation structure.

In this section, we focus on specifying models for trend and serial correlation separately. The first-step procedure involves modeling the trend component, whereas the second-step procedure focuses on modeling the stationary and non-stationary processes, which capture the serial correlation. The first-step procedure is necessary because COVID-19 data typically do not follow a linear deterministic time trend. Instead, the mode of each observation is represented as a function of the time index. The empirical analysis in Section 5.1 demonstrates that the nonlinearity of the COVID-19 trend is more complex than a simple linear trend process. The second-step procedure aims to transform the COVID-19 data into a stationary process with serial correlation. By accounting for the serial correlation, we can capture the temporal dependencies present in the data and ensure that modal regression analysis is applicable. By separately specifying models for trend and serial correlation, we can effectively capture the characteristics of time-series data and enhance the accuracy of the modal regression analysis of COVID-19 data.

In the first-step procedure, we focus on specifying the nonlinear trend component. We employ three estimation methods. The first is the B-spline modal regression (BMR) proposed by Yu et al. (2020). To implement the BMR, we utilize B-splines, which are a type of basis function. We define ξ as the set of the knots of the expected B-spline that partitions the unit interval $[0, 1]$. Each $t_i \in [0, 1]$ corresponds to a knot, and we construct the linear spline space using $(k - \ell)$ B-spline basis functions of order ℓ. Here, the unit interval represents the time index space obtained by dividing the time by the sample size n; that is,$\{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, \frac{n}{n}\}$. The basis functions are defined as follows: for $i = 0, 1, \ldots, k - \ell - 1$,

$$B_{i,0}(\tau) := \begin{cases} 1, & \text{if } t_i \leq \tau < t_{i+1} \\ 0, & \text{otherwise,} \end{cases}$$

[3]It is also possible to interpret this association in the opposite manner: for a fixed κ, if $g(X) = f(X)$, the interval around the conditional mode can cover more samples than any other quantity $g(X) \neq f(X)$. This implies that τ increases as $g(X)$ approaches $f(X)$.

and

$$B_{i,\ell}(\tau) := \frac{x - t_i}{t_{i+\ell} - t_i} B_{i,i-\ell}(\tau) + \frac{t_{i+\ell+1} - x}{t_{i+\ell+1} - t_{i+1}} B_{i+1,\ell-1}(\tau),$$

where $B_{i,\ell}(\cdot)$ represents the i-th basis function of order ℓ. For simplicity, we denote $B_{i,\ell}(\cdot)$ as $B_i(\cdot)$, and the B-spline basis function is denoted as $B(\cdot) = B_0(\cdot), B_1(\cdot), \ldots, B_{(k-\ell-1)}(\cdot)$. In this setup, we assume that the conditional mode is a linear function of $B(\tau)$, that is, $m(\tau, \theta) = \theta' B(\tau)$, and estimate the parameter vector $\theta = (\theta_0, \ldots, \theta_{k-\ell-1*})'$ using modal regression. Note that the choice of (ξ, ℓ) has a deterministic effect on the shape of the spline function. The selection of ξ determines the positioning of control points, whereas ℓ determines the number of coefficients in each piece of the piecewise polynomial representation.

Different types of B-splines can be defined by choosing (ξ, ℓ) in various ways. Examples of well-known B-splines are as follows. First, for the nonperiodic B-spline, the first and last m knots are fixed at 0 and 1, respectively, where $2m < k + 1$. For instance, if we choose $\xi = \{0, 0, 0, 0.3, 0.6, 1, 1, 1\}$, it corresponds to the nonperiodic B-spline knots. Second, the uniform B-spline assumes equally spaced knots. For example, if we select $\xi = \{0, 0.25, 0.5, 0.75, 1\}$, it represents the uniform B-spline knots. Third, Bezier knot lets $k = 1$ such that the knots are set as $\xi = \{0, 1\}$.

The choice of (ξ, ℓ) is typically determined empirically based on the data characteristics. For example, if the data exhibit a curve resembling a quadratic function, the non-periodic B-spline can be employed by setting the first and last $\ell+1$ knots to 0 and 1, respectively, and placing ℓ knots in the middle. Yu et al. (2020) demonstrate that the BMR is robust against outliers or heavy-tailed error distributions. Moreover, they show that BMR performs no worse than least squares estimation when the errors are normally distributed.

Second, the local polynomial modal regression (LPMR) method, as examined by Xiang and Yao (2022), can be employed as a second estimation method to estimate the modal function. The LPMR involves the application of a p-th order Taylor expansion for the conditional mode around a reference value of $\tau \in [0, 1]$, specifically $\frac{t}{n}$, to approximate $f(\tau)$. This approximation can be expressed as

$$f(\tau) \approx \sum_{i=0}^{p} \frac{f^{(i)}\left(\frac{t}{n}\right)}{i!} \left(\tau - \frac{t}{n}\right)^i = \sum_{i=0}^{p} \theta_{i*} \left(\tau - \frac{t}{n}\right)^i,$$

where $\theta_{i*} := f^{(i)}\left(\frac{t}{n}\right)/i!$, and $f^{(i)}(\cdot)$ represents the i-th order derivative of $f(\cdot)$. Similar to the B-spline modal regression, the parameter vector θ_* can be estimated by applying modal regression. In the LPMR method, the optimal order p can be selected by minimizing the CQF with respect to the degree of the polynomial. Xiang and Yao (2022) demonstrate through simulation that the LPMR method complements the conventional nonparametric mean and median regressions, particularly in the presence of outliers. Further-

more, they show that LPMR exhibits better prediction performance for skewed data than mean and median regressions.

Finally, we apply the linear modal regression (MR) using the approach described in Yao and Li (2014). This method assumes a linear model for $f(\frac{t}{n})$ as $m(\frac{t}{n}, \theta_*) = \theta_0 + \theta_{1*}\frac{t}{n}$, where $\theta_* := (\theta_0, \theta_{1*})'$ is estimated through modal regression. Importantly, this linear model is not suitable for representing COVID-19 data. However, we utilize it as a benchmark model to contrast its linearity against a process with a linear trend.

In addition to these methods, Zhou and Huang (2019) propose the mean shift modal regression (MSMR) method for nonparametric trend prediction. However, the forecasting error of MSMR is greater than that of the first two methods (BMR and LPMR). As a result, we focus on the BMR, LPMR, and MR methods to predict the trends in our analysis.

As the second step in specifying a correlation model, we apply an autoregressive (AR) model. The AR model is defined as follows:

$$Y_t = \alpha_* + \sum_{i=1}^{\ell} \beta_{i*} Y_{t-i} + U_t, \tag{1}$$

where α_* and β_{i*} are the estimated parameters and U_t is the error term. We distinguish between the two versions of the AR model based on the type of error term. First, we use a mean autoregressive (MEAR) process that assumes that U_t follows a white noise process with zero mean and constant variance. This is the conventional AR process. However, for modal regression, we assume that U_t is a white noise process with a zero conditional mode, given $\mathcal{F}_t$, where $\mathcal{F}_t$ is the sigma-algebra generated by Y_{t-1}, Y_{t-2}, and so on. We refer to this version as the modal autoregressive (MAR) process.

If the series Y_t is not stationary, we apply differencing to obtain a stationary process. In this case, we replace Y_t and Y_{t-i} in (1) with ΔY_t and ΔY_{t-i}, respectively, and estimate the parameters $\theta_* := (\alpha_*, \beta_{1*}, \ldots, \beta_{\ell*})'$ using modal regression. If ΔY_t is still nonstationary, we repeat the differencing process until we obtain a stationary process. The MAR model for a differenced process of order d can be expressed as:

$$\Delta^d Y_t = \alpha_* + \sum_{i=1}^{\ell} \beta_{i*} \Delta^d Y_{t-i} + U_t.$$

The lag order ℓ is determined by applying the Bayesian information criterion (BIC), which is commonly used to estimate the MEAR process. As U_t is determined by the conditional distribution of $\Delta^d Y_t$, given $\Delta^d Y_{t-i}$ $(i = 1, 2, \ldots)$, the BIC can effectively assist in estimating the lag order of the MAR process.

Before moving to the next section, we note that the forecasting methodology by modal regression for integrated series has not yet been fully developed in the literature. The methodology has to be developed

by estimating the cointegrating relationship between the integrated series, that distinguishes between the long-run and short-run relationships. Nevertheless, the current literature does not provide a methodology for this to our best knowledge. Due to this fact, we limit our interest to forecasting the future observations by the lagged dependent variables and leave developing the methodology as a future research topic.

4 Evaluation of the Modal Regression by Simulation

In this section, we perform Monte Carlo simulations to examine the application of modal regression and evaluate its effectiveness compared with other estimation methods. We consider two types of data generating processes (DGPs): cross-sectional and time-series data. Through these Monte Carlo simulations, we aim to gain a comprehensive understanding of the capabilities and limitations of modal regression, particularly in comparison with alternative estimation approaches, in both cross-sectional and time-series settings.

4.1 Simulation Using Cross-Sectional Data

We demonstrate the application of modal regression by comparing it with mean and median regressions. First, we generate a set of identically and independently distributed (IID) observations $\{(X_t, Y_t) : t = 1, \ldots, n\}$ according to the following formula: $Y_t = \alpha_* + \beta_* X_t + U_t$. In this case, we set $\alpha_* = 0$, $\beta_* = 2$, $X_t \sim_{iid} U[0, 1]$, and $U_t \sim_{iid} 0.5N(-2, 3^2) + 0.5N(2, 1^2)$ to ensure that X_t and U_t are independent. We refer to this data generating process as DGP1. The left panel of Figure 1 displays the density function of U_t, where we observe the expected value $\mathbb{E}(U_t) = 0$, median $\text{median}(U_t) = 1$, and mode $\text{mode}(U_t) = 2$. Consequently, the following relationships hold:

$$\mathbb{E}\left(Y_t \mid X_t\right) = \alpha_{1*} + \beta_{1*} X_t = 2X_t, \quad \text{median}\left(Y_t \mid X_t\right) = \alpha_{2*} + \beta_{2*} X_t = 1 + 2X_t, \quad \text{and}$$

$$\text{mode}\left(Y_t \mid X_t\right) = \alpha_{3*} + \beta_{3*} X_t = 2 + 2X_t.$$

This implies that the conditional mean, median, and mode functions are associated with different parameter values. The left panel of Figure 2 illustrates the three functions represented by the blue, orange, and red lines, respectively. Additionally, 200 observations randomly drawn from DGP1 are presented. We can see that more observations align with the red line, indicating an asymmetric conditional distribution. The forecast band encompasses a higher concentration of observations around the conditional mode function than around the other functions. To estimate these three functions, we utilize the mean regression (MER), linear quantile regression (LQR) with a quantile level of 0.5, and MR, respectively.

<Insert Figures 1 and 2 around here>

As highlighted by Ullah et al. (2022), the choice of bandwidth is crucial for the estimation results when using the MR method. In our simulations, we explore different bandwidth values and observe that the results are comparable when the SJ's, Scott's, and Silverman's bandwidth selection methods are used. Therefore, we discuss the simulation results by focusing on the three density function estimations. We use a two-step approach to estimate the density function. First, we employ a least squares estimation to obtain conditional residuals. Subsequently, we optimize the conditional density function using the MEM algorithm, as described in Section 3.1. This enables us to obtain reliable estimates of the density function.

We present the simulation results in the first panel of Table 2. We conduct 1,000 independent experiments for different sample sizes ($n = 100$, 200, 300, 500, and 1,000) and report the MSEs of the estimated coefficients. The MSE provides a measure of the average squared difference between the estimated and true coefficients across the simulation experiments.

<Insert Table 2 around here>

The simulation results are summarized as follows:

(a) The MER, LQR, and MR methods consistently estimate the unknown parameters. As the sample size n increases, the MSEs decrease for all three methods, indicating an improved estimation accuracy with larger sample sizes.

(b) The MSEs obtained by the MR method are generally similar among the different density function estimation methods. However, a regular rank relationship is observed among the MR methods. In particular, the MSEs obtained using SJ's bandwidth tend to be smaller than those obtained using the other methods. However, there is no consistent rank relationship between the MSEs obtained using Scott's and Silverman's bandwidths. This suggests that SJ's bandwidth tends to provide more accurate estimation results in terms of MSE for the MR method, whereas the relative performance of Scott's and Silverman's bandwidths may vary depending on the specific data and model conditions.

(c) We also examine the distribution of the estimated coefficients. The upper panels of Figure 3 show the estimated probability density functions of $\widehat{\alpha}_n$ and $\widehat{\beta}_n$ obtained by the MER method using SJ's bandwidth. It can be observed that the empirical distributions of the estimated coefficients are close to a bell-shaped distribution. This indicates that the estimated coefficients tend to be centered around their true values and the variability around the mean is relatively symmetric. □

<Insert Figure 3 around here>

Using DGP1, we verify that the MEM algorithm consistently estimates the conditional mode, whereas the MER and LQR estimations estimate the conditional mean and median, respectively.

We further explore the modal regression by considering a different DGP condition, DGP2. In DGP2, we assume another distribution for the error term instead of the mixture normal distribution. Similar to DGP1, we have $Y_t = \alpha_* + \beta_* X_t + U_t$, where $\alpha_* = 0$, $\beta_* = 1$, $X_t \sim_{iid} U[0,1]$, and $U_t \sim_{iid} \mathcal{X}_3^2 - 3$ with $X_t \perp U_t$, indicating that U_t follows a chi-squared distribution with three degrees of freedom shifted by three. The right panel of Figure 1 shows the density function of U_t for DGP2. We specifically examine DGP2 to investigate the performance of the modal regression when the error distribution has fat tails. The density function of U_t exhibits an extreme left fat tail owing to the truncation at the border from the left, and the right tail is fatter than that of a normal distribution by construction. Consequently, we anticipate that the modal regression will perform relatively poorly compared to DGP1, as more observations are required to estimate the density function accurately. Note that $\mathbb{E}(U_t) = 0$, median$(U_t) = -0.63$, and mode$(U_t) = -2$. Consequently, we have

$$\mathbb{E}\left(Y_t \mid X_t\right) = \alpha_{4*} + \beta_{4*} X_t = X_t, \quad \text{median}\left(Y_t \mid X_t\right) = \alpha_{5*} + \beta_{5*} X_t = -0.63 + X_t, \quad \text{and}$$

$$\text{mode}\left(Y_t \mid X_t\right) = \alpha_{6*} + \beta_{6*} X_t = -2 + X_t.$$

The right panel of Figure 2 displays the three different functions along with 200 observations randomly drawn from DGP2. Similar to DGP1, more observations are distributed along the conditional mode function.

Using the observations generated from DGP2, we perform independent experiments following the same procedure as that for DGP1, and the results are presented in the second panel of Table 2. The simulation results are summarized as follows:

(a) As for DGP1, the MER, LQR, and MR methods consistently estimate the unknown parameters for DGP2. Similar to DGP1, we observe that the MSEs decrease as the sample size n increases for all three estimation methods. While the MSEs obtained using the MR methods are generally larger than those obtained using the MER and LQR methods, we can confirm that the MR methods provide consistent MR estimators. For brevity, we do not report the detailed simulation results. However, notably, as the sample size increases to 2,000, the MR estimators become very close to the true unknown parameters.

(b) When comparing the MSEs obtained by the MR methods for DGP2, we find that they are similar in general. However, this is consistent with the observations for DGP1 that the MSE obtained by SJ's method is smaller than those obtained by Scott's and Silverman's methods. This suggests that SJ's

bandwidth selection method tends to yield more accurate results in terms of MSE than the other two methods.

(c) The lower panels of Figure 3 depict the empirical density functions of $\widehat{\alpha}_n$ and $\widehat{\beta}_n$ obtained by the MER method using SJ's bandwidth for DGP2. While the empirical distributions are not perfectly bell-shaped, they gradually converge to a normal distribution. However, this convergence is slower than that in DGP1 owing to the presence of fat tails in the error distribution. The fat tails contribute to deviations from perfect normality in the empirical density functions. □

From the additional simulation using DGP2, we observe that the modal regression effectively estimates the conditional mode. Despite the presence of fat tails in the error distribution, estimating the density function using SJ's bandwidth remains more efficient than using the other methods. This result suggests that SJ's bandwidth selection is robust and effective in capturing the characteristics of conditional mode, even in the presence of non-normal and fat-tailed error distributions.

In addition to the simulations presented in this study, other simulations were conducted by assuming different error distributions and consistent results were obtained. For instance, when considering an asymmetric beta distribution for U_t, the parameter estimators obtained through modal regression exhibited faster convergence to the unknown parameters. This further supports the effectiveness and robustness of the modal regression under various error distribution scenarios.

4.2 Simulation Using Time Series Data

We further extend our simulation by considering serially correlated time-series data. As an extreme case, we first examine the unit-root process given by

$$Y_t = \alpha_* + \beta_* Y_{t-1} + U_t,$$

where $\alpha_* = 0.3$, $\beta_* = 1$, and $U_t \sim_{\text{iid}} \Gamma(3,1) - 3$ with $Y_{t-1} \perp U_t$, indicating that U_t follows a gamma distribution. Note that $\mathbb{E}(U_t) = 0$, median$(U_t) = -0.3$, and mode$(U_t) = -1$. Consequently, the following equations are derived:

$$\mathbb{E}\left(Y_t \mid Y_{t-1}\right) = \alpha_{7*} + \beta_{7*} Y_{t-1} = -0.3 + Y_{t-1}, \quad \text{median}\left(Y_t \mid Y_{t-1}\right) = \alpha_{8*} + \beta_{8*} Y_{t-1} = Y_{t-1}, \quad \text{and}$$

$$\text{mode}\left(Y_t \mid Y_{t-1}\right) = \alpha_{9*} + \beta_{9*} Y_{t-1} = -0.7 + Y_{t-1}.$$

We conduct simulations to examine the behavior of the estimators in terms of MSE. The experimental results are presented in the first panel of Table 3, obtained by performing 1,000 independent experiments. Similar to previous simulations, we estimate the bandwidth components used in the density function estimation by first estimating the conditional mean. The simulation results are summarized as follows:

(a) For all the cases considered in our simulations, we observe a consistent decrease in MSE as the sample size n increases. This indicates that the MER, LQR, and MR methods are consistent in estimating the unknown parameters. Moreover, the decreasing trend of MSEs as n increases suggests that modal regression is an effective estimation method, even for the unit-root process.

(b) When comparing the MSEs obtained by the MR methods, we consistently observe that the MSE obtained by SJ's method is overall smaller than those obtained by the other two methods, as we have previously observed in the cross-sectional data simulations. This suggests that SJ's method performs better in terms of MSE when estimating the density function in the modal regression framework, regardless of the data generating process.

(c) Figure 4 displays the estimated probability density functions using $\widehat{\alpha}_n$ and $\widehat{\beta}_n$ obtained by the MAR method using SJ's bandwidth. The figure shows that the empirical density functions exhibit shapes that are close to a normal distribution. However, a more thorough investigation is required to fully understand the influence of the unit-root process on the asymptotic distribution. In particular, because the mean and median regressions do not produce asymptotically normally distributed estimators, it would be interesting to examine the behavior of the asymptotic distribution under the modal regression framework. Although the current study does not conduct a detailed analysis, the empirical density functions obtained from the simulation results appear to resemble a bell-shaped distribution as the sample size increases. Investigating the asymptotic distribution in the presence of a unit-root process is a promising avenue for future research. This analysis would shed light on the behavior and properties of the estimators obtained through the modal regression, providing a more comprehensive understanding of their performance in the context of unit-root processes. □

<Insert Table 3 and Figure 4 around here>

We extend our investigation of modal regression by examining an alternative DGP condition for stationary time-series data, namely, DGP4. In DGP4, we assume stationary series instead of a unit-root process. Similar to DGP3, our model is defined as $Y_t = \alpha_* + \beta_* Y_{t-1} + U_t$, where $\alpha_* = 0.3$, $\beta_* = 0.5$, and $U_t \sim_{iid} \Gamma(3,1) - 3$ with $Y_{t-1} \perp U_t$. We conduct a specific examination of DGP4 to investigate the perfor-

mance of modal regression using stationary data. Consequently, we have

$$\mathbb{E}\left(Y_t \mid Y_{t-1}\right) = \alpha_{10*} + \beta_{10*} Y_{t-1} = -0.3 + 0.5 Y_{t-1}, \text{median}\left(Y_t \mid Y_{t-1}\right) = \alpha_{11*} + \beta_{11*} Y_{t-1} = 0.5 Y_{t-1}, \text{and}$$

$$\text{mode}\left(Y_t \mid Y_{t-1}\right) = \alpha_{12*} + \beta_{12*} Y_{t-1} = -0.7 + 0.5 Y_{t-1}.$$

Using the observations generated from DGP4, we perform independent experiments following the same procedure as that for DGP3, and the results are presented in the second panel of Table 3. The simulation results are summarized as follows:

(a) Regarding DGP4, the MER, LQR, and MR methods consistently estimate the unknown parameters. Similar to DGP3, we observe a decrease in the MSEs as the sample size n increases for all three estimation methods.

(b) When comparing the MSEs obtained by the MR methods for DGP4, we observe a general similarity, aligning with the findings for DGP3. That is, the MSE obtained through SJ's method is smaller than those obtained through Scott's and Silverman's methods. This suggests that SJ's bandwidth selection method tends to outperform the other two alternatives, even in stationary series.

(c) The lower panels of Figure 4 depict the empirical density functions of $\widehat{\alpha}_n$ and $\widehat{\beta}_n$ obtained by the MR method using SJ's bandwidth for DGP4. The empirical distributions are also bell-shaped, and they gradually converge to normal distributions. □

In the additional simulation employing DGP4, it becomes evident that modal regression effectively estimates the conditional mode for time-series data. This outcome implies that SJ's bandwidth selection is robust and proficient in capturing the characteristics of the conditional mode in dealing with stationary or non-stationary data.

5 Empirical Analysis

In this section, we analyze the trends in COVID-19 processes in Canada, Japan, South Korea, and the U.S. using the modal regression approach to forecast the COVID-19 confirmed cases. To conduct this analysis, we utilize the official data on COVID-19 provided by the Johns Hopkins University Center for Systems Science and Engineering (JHU-CSSE). This data collection incorporates information from various sources, such as the WHO, national governments, and local media reports, enabling comprehensive tracking of the disease. Starting from January 22, 2020, JHU-CSSE has been updating and publishing daily data on the cumulative number of confirmed COVID-19 cases, deaths, and recoveries for each country and territory.

These datasets serve as the basis for our analysis, allowing us to examine the patterns and dynamics of the COVID-19 pandemic in Canada, Japan, South Korea, and the U.S. from the modal regression perspective.[4] The goal of this section is in forecasting the confirmed cases. As we detail below, the processes of the confirmed cases are not stationary, exhibiting trends as Figure 5 demonstrates.

We are particularly interested in forecasting the future confirmed cases by basing on their past values and by modal regression. As we mentioned above, the cointegration methodology for forecasting the confirmed cases by associating them with other integrated series via modal regression is not yet fully developed in the literature. We, therefore, limit our interest to forecasting the future confirmed cases based on their past values.

5.1 Forecasting the Confirmed COVID-19 Cases by Trend Fitting

Using the daily data for cumulative confirmed COVID-19 cases from February 8, 2022, to April 8, 2022, we apply various nonlinear trend estimation methods, including BMR, LPMR, and MR, in addition to the MER and LQR methods. Each country's dataset consists of 60 observations, which we split into two parts. The first 50 observations are used as the training set for trend estimation to represent our in-sample measures. The remaining 10 observations serve as the test set for evaluating the performance of each method, reflecting our out-of-sample measures. To assess the performance of each method, we use three evaluation metrics: RMSE, MAE, and CQF. These metrics provide measures of accuracy and reliability for comparing the estimated trends across different methods and countries.

In the empirical applications, we follow a three-step process. First, min-max normalization was applied to standardize the data. The transformed data are denoted as Y_t' and are calculated as

$$Y_t' := \left(\frac{Y_t - Y_0}{Y^0 - Y_0} \right),$$

where $Y^0 := \max_{t=1,2,\ldots,n} Y_t$ and $Y_0 := \min_{t=1,2,\ldots,n} Y_t$. This normalization converts a large number of confirmed cases to a range between zero and one. Additionally, we adjust the time index t to be within the unit interval by rescaling it to $\frac{t}{n}$. Consequently, Y_t' is readjusted to Y_τ', where $\tau = \frac{1}{n}, \frac{2}{n}, \ldots, 1$. This adjustment allows us to fit the trend of $Y_{(\cdot)}'$. In the second step, we assume that $Y_{(\cdot)}'$ follows a linear trend and estimate the intercept and linear coefficient using the MER, LQR, and MR methods. Subsequently, we compute the RMSE, MAE, and CQF using both the training and test sets. Finally, we consider the possibility of $Y_{(\cdot)}'$ having a nonlinear trend and estimate the trend using the BMR and LPMR methods.

[4]The data are available at the following URL: https://github.com/CSSEGISandData/COVID-19

For the BMR method, we set the B-spline order ℓ to 3, following Yu et al. (2020), and select 11 knots, denoted as $\xi = \{0, 0, 0, 0, 0.2, 0.4, 0.6, 1, 1, 1, 1\}$. These knots are selected based on the observation that the empirical curves resemble quadratic functions. For the LPMR method, we determine the polynomial degree by minimizing the CQF. We also compute the RMSE, MAE, and CQF using the training and test sets, similar to the linear trend case.

<Insert Table 4 and Figure 5 around here>

We present the estimation and prediction results in Figure 5 and Table 4. Figure 5 shows the scatter plots of all data observations considered, along with the regression lines obtained from the different regression methods. To convert the predicted values $Y'_{(\cdot)}$ back to the original scale, we use the formula $\widehat{Y}_{(\cdot)} := \widehat{Y}'_{(\cdot)}(Y^0 - Y_0) + Y_0$, where $\widehat{Y}'_{(\cdot)}$ represents the series predicted by each regression method. Table 4 presents the RMSE, MAE, and CQF for the four countries. Based on the results, we summarize the estimation and forecast results as follows:

(a) The MER, LQR, and MR methods demonstrate superior performance in terms of RMSE, MAE, and CQF, respectively. This finding aligns with the characteristics of these estimators and their relationships with the respective objective functions.

(b) In the training set, the modal regression methods, particularly the LPMR and BMR methods, exhibit superior performance compared with the mean and median regression methods for all four countries. In the test set, the LPMR and BMR methods continue to outperform the other regression methods for Japan and the U.S., whereas the MR method outperforms the other regression methods for Canada and South Korea. This trend is illustrated in Figure 5. Moreover, the LPMR method consistently outperforms the other modal regression methods. The optimal value of p is chosen by minimizing the CQF. Additionally, Table 4 illustrates that the LPMR method achieves lower RMSE and MAE values than the BMR method in most cases. In terms of CQF, both the LPMR and BMR methods offer competitive performance.

(c) Although estimating a linear modal trend is straightforward, its performance, as measured by the CQF, is consistently surpassed by the BMR and LPMR methods. This indicates that the COVID-19 confirmed cases in the four countries do not conform to a linear trend.

(d) For the test sets of South Korea and the U.S., the forecast lines generated by the MR method are closer to the actual values compared with those produced by the MER and LQR methods. This indicates that the MR method performs better in forecasting the highest and lowest values of confirmed cases in South Korea and the U.S., respectively. This implies that if the South Korean government relies

solely on the mean and median regression perspectives to formulate health policies, there is a risk that healthcare systems will be overwhelmed by the actual number of confirmed cases, surpassing the forecasts based on mean and median regressions. Similarly, implementing health policies in the U.S. without considering the conditional mode forecast may lead to the misallocation of financial and human resources. □

By comparing the different regression methods, we find that modal regression outperforms mean and median regressions in both the training and test sets for all four countries, especially in terms of the CQF. This observation confirms that the modal regression methods, which emphasize the conditional mode, yield narrower forecast intervals than traditional approaches that primarily focus on the characteristics of the conditional mean and median to achieve better RMSE or MAE scores, respectively. The superior performance of modal regression suggests that considering the conditional mode can lead to more accurate and precise predictions in the context of COVID-19 trend analysis.

5.2 Forecasting the Confirmed COVID-19 Cases by Modal Autoregression

Using the MAR method, we analyze the transformed COVID-19 cases $Y'_{(\cdot)}$ from Canada, Japan, South Korea, and the U.S. for forecasting purposes. Before applying the MAR method, we perform the augmented Dickey-Fuller (ADF) test to examine the presence of a unit root in the series. If the unit-root hypothesis cannot be rejected, we take the first difference of the series and repeat the ADF test. This procedure is repeated until we obtain evidence rejecting the unit-root hypothesis. This iterative process allows us to identify the appropriate order of differencing required for the time-series data. By ensuring stationarity in the data, we can effectively apply the MAR method to forecast COVID-19 cases.

Table 5 presents the results of the ADF test conducted on the original series and the series that has undergone three levels of differencing. The p-values obtained from the ADF test for the original series are all greater than 1%, indicating that we cannot reject the null hypothesis of a unit root for these series. However, for the three-times differenced series, the p-values are all less than 1%, providing evidence to reject the unit-root hypothesis. Based on these results, we specify the MAR model for the three-times differenced data and estimate the unknown parameters using the modal regression approach. Table A.1 in the Appendix provides the descriptive statistics of the three-times differenced data for each country. The MAR order is determined by minimizing the BIC, as described in Section 3.2. In this case, we find that a MAR order of 6 is selected for all four countries.

<Insert Table 5 around here>

We assess the performance of the forecasts obtained from the MAR model estimation by comparing them with the forecasts obtained from the MEAR and LQAR models. To evaluate the forecasts on the training data, we compare the forecasts with the actual observations and compute the RMSE, MAE, and CQF. We employ two approaches to evaluate the forecasts on the test data. First, we forecast the future values sequentially, using the recent forecasts as inputs (non-teacher-forcing method). Second, we forecast the future values by utilizing the recent realizations from the test data as inputs (teacher-forcing method). The teacher-forcing method is expected to produce more accurate forecasts than the non-teacher-forcing method because forecast errors tend to accumulate over the forecasting period. In the context of COVID-19 forecasting, this can lead to compounded errors over time. For instance, if the model underestimates cases in one period, this underestimation carries forward and can result in a consistently lower projection in subsequent periods. However, in situations where we may not have access to accurate future COVID-19 observational data, the non-teacher-forcing method holds a distinct advantage. By employing these evaluation methods, we can assess the precision and accuracy of the MAR model forecasts and compare them with those of the MEAR and LQAR models.

We present the qualitative forecasting results obtained from the MAR model using the teacher-forcing method. The four upper panels of Figure 6 show the results. The MAR coefficients are estimated using modal regression with the bandwidth selected as SJ's bandwidth. It is evident that the MEAR, LQAR, and MAR models exhibit impressive forecasting performance in capturing the trend of the cumulative confirmed COVID-19 cases. Furthermore, the lower four panels of Figure 6 show the forecasting results obtained using the non-teacher-forcing method. These results demonstrate a pattern similar to that of the forecasting results shown in the upper panels.

<Insert Figure 6 around here>

We provide a visual comparison of the daily forecasts of the cumulative confirmed COVID-19 cases in Figure 6. It is important to note that differencing the cumulative confirmed cases results in the forecast of daily confirmed cases. To visualize the daily forecast, we present Figure 7, which displays the daily forecast for the four countries. The results are summarized as follows:

(a) For both the teacher-forcing and non-teacher-forcing methods, we observe that the MAR method accurately captures the trends and autocorrelation patterns in the daily confirmed cases. Specifically, for the U.S., the series exhibits a declining trend with oscillation, whereas the series for Japan shows an initial decline followed by a rebound. These patterns are accurately captured by the MAR model, demonstrating its superior performance compared with the MEAR and LQAR models.

(b) For both the teacher-forcing and non-teacher-forcing methods, we observe that the MAR method outperforms the MAR and MEAR models in forecasting outliers. In Figure 7, the MAR forecasts show wider variations than the LQAR and MEAR forecasts. This indicates that the MAR forecasts are better able to capture the cyclical peaks and bottoms of the daily confirmed cases. This trend is particularly evident in Canada and Japan.

(c) The forecast error on the test set tends to be larger in the non-teacher-forcing method compared to the teacher-forcing method. This is because the prediction errors in the non-teacher-forcing method accumulate as the last-period forecast is used as a covariate for the next-period forecast. This can lead to compounding errors and potentially larger forecasting errors over time. However, in the teacher-forcing method, the use of actual realizations in the test data as inputs for forecasting helps mitigate the accumulation of errors, resulting in generally more accurate forecasts.

(d) The forecast results obtained from the MAR, LQAR, and MEAR models show that governments can be better prepared when forecasting daily confirmed cases using the MAR model compared to the LQAR and MEAR models. The MAR model demonstrates better performance in capturing the trends, autocorrelation patterns, and outliers in the daily confirmed cases. This implies that the MAR model for forecasting can provide governments with more accurate and reliable information to make informed decisions and take appropriate measures in response to the COVID-19 pandemic. □

<Insert Figure 7 around here>

The qualitative prediction outcomes for daily confirmed cases using the MAR model estimated by modal regression with Scott's and Silverman's bandwidths are shown in Figures 8 and 9. They exhibit similar performance to the forecasts obtained using SJ's bandwidth. However, there are significant differences in the quantitative evaluations.

<Insert Table 6 and Figures 8 and 9 around here>

The quantitative results using the three different bandwidths are presented in Tables 6 and 7. Table 6 reports the RMSEs between the forecasts and the actual values obtained using the teacher-forcing method. In addition, Table 6 displays the estimated parameters obtained using SJ's, Scott's, and Silverman's bandwidths for the MAR model, whereas Table 7 presents the performance measures of the MEAR, LQAR, and MAR models, including the RMSE, MAE, and CQF obtained by the teacher-forcing and non-teacher-forcing methods. Here, SJ's bandwidth is used for the MAR models. The quantitative results are summarized as follows:

(a) Table 6 shows that SJ's bandwidth outperforms the other two methods (Scott's and Silverman's) in terms of RMSE for Canada, Japan, South Korea, and the U.S., which is consistent with the findings in Section 4. The estimated coefficients are generally similar across the different bandwidths for Canada, South Korea, and the U.S., whereas Japan exhibits notable differences among the different bandwidths.

(b) Table 7 indicates that for the training sets of the four countries, the CQF for the MAR model consistently has the smallest value. For the test set and for both teacher-forcing and non-teacher-forcing methods, the MAR model most often performs better than the MEAR and LQAR models in terms of the CQF. Although we do not report here, the same result is also obtained when applying Scott's and Silverman's bandwidths, and this can be easily inferred from the forecast results in Figures 7, 8, and 9. Furthermore, the MAR model sometimes achieves even the best RMSE and MAE. This can be attributed to the robustness of the modal regression approach in handling unexpected noise and outliers in the differenced data.

(c) The results obtained by the teacher-forcing method consistently outperform the forecast obtained by the non-teacher-forcing method. This validates our earlier discussion on the difference between these two methods, highlighting that the non-teacher-forcing method tends to accumulate forecast errors over the forecast horizon, leading to less accurate predictions than the teacher-forcing method. □

<Insert Table 7 around here>

When comparing the performances of the different bandwidths for the four countries, it is evident that SJ's bandwidth consistently yields smaller RMSE values, indicating its superiority in forecasting confirmed COVID-19 cases using the MAR method. Additionally, the smaller CQF values obtained by the MAR method compared with those obtained by the MEAR and LQAR methods demonstrate that the MAER and LQAR methods are more sensitive to outliers and tend to accumulate forecasting errors gradually over time, thereby affecting the overall forecast accuracy.

6 Conclusion

Since the outbreak of COVID-19, there has been a growing interest in analyzing its trend in the scientific literature. Over the years, our understanding of the disease and its development has improved as we have accumulated more data and gained more experience. In this context, the modal regression method has emerged as a valuable statistical tool for handling noisy and skewed data to predict and analyze the COVID-

19 trend. This allows us to account for the characteristics and fluctuations in the data, leading to more accurate predictions and insights into the dynamics of the disease.

This study analyzes the confirmed COVID-19 cases using the modal regression approach, which involves four main steps. First, an objective function is formulated to evaluate the different forecasts for a series, considering the estimation of conditional mean, median, and mode. The forecasts are evaluated based on RMSE, MAE, and CQF metrics. Modal regression aims to optimize the CQF, whereas conditional mean and median regressions optimize the RMSE and MAE, respectively. The CQF measures the probability of a random variable falling within a forecasted interval, making the conditional mode function suitable for optimizing the CQF. Second, the prediction models available in the literature are reviewed, focusing on their application in modal regression. Two types of models for time-trend and unit-root processes are examined for their suitability in the modal regression framework. Third, simulations are performed to investigate the properties of the modal regression. The simulations involve cross-sectional, stationary, and unit-root data, and the consistency of the modal regression is examined. It is discovered that the performance of the modal regression critically depends on the choice of bandwidth used to estimate the density function. Notably, the results demonstrate that SJ's bandwidth, along with Scott's and Silverman's bandwidths, provides robust estimation outcomes. Finally, the modal regression approach is applied to analyze the confirmed COVID-19 cases in Canada, Japan, South Korea, and the U.S. The empirical analysis aims to forecast and analyze the trends in COVID-19 cases using the modal regression framework, considering the characteristics and dynamics of the data for these countries.

Several key findings emerged from our empirical analysis. First, the MR method consistently outperforms the other methods in terms of the CQF for the cumulative confirmed COVID-19 case data, regardless of whether time-trend or unit-root models are considered. For all four countries (Canada, Japan, South Korea, and the U.S.), the CQF achieved through modal regression is consistently smaller than that of the other methods, indicating better performance in terms of capturing the forecast uncertainty. Second, the forecasts obtained through modal regression demonstrate superior capability in capturing the cyclical peaks and bottoms of daily confirmed COVID-19 cases compared with those obtained through mean and median regressions. This is attributed to the wider variation in the modal regression forecasts, which align more closely with the actual cyclical patterns. In contrast, the forecasts from the mean and median regressions tend to underestimate and overestimate the cyclical peaks and bottoms, respectively. This finding has important economic implications, suggesting that modal forecasting can help governments avoid high risks when formulating health policies related to COVID-19.

The research methodology employed in this study can be extended to analyze data from other countries

or explore other infectious diseases. The methodology does not assume specific characteristics of the data from the four countries considered in this study and is established based on methodological considerations related to modal regression. Therefore, it can be used to analyze various time-series data.

However, we also emphasize that the current study focuses on forecasting the future observations of a single series using its past observations when it is nonstationary. It does not examine the interrelationship between two or more nonstationary processes. For instance, in the case of nonstationary data such as confirmed COVID-19 cases, it would be invaluable to investigate the cointegrating relationships with other variables in the quantile autoregressive distributed lag model framework (e.g., Cho, Kim, and Shin, 2015). Exploring these interrelationships could be a potential direction for future research.

References

AJIDE, K. B., R. L. IBRAHIM, AND O. Y. ALIMI (2020): "Estimating the impacts of lockdown on Covid-19 cases in Nigeria," *Transportation Research Interdisciplinary Perspectives*, 7, 100217.

ALMESHAL, A. M., A. I. ALMAZROUEE, M. R. ALENIZI, AND S. N. ALHAJERI (2020): "Forecasting the spread of COVID-19 in Kuwait using compartmental and logistic regression models," *Applied Sciences*, 10, 3402.

ALMOND, D., X. DU, AND S. ZHANG (2020): *Did COVID-19 improve air quality near Hubei?*, National Bureau of Economic Research Cambridge, Massachusetts.

ANDRÉE, B. P. J. (2020): "Incidence of COVID-19 and connections with air pollution exposure: evidence from the Netherlands," *MedRxiv*, 2020–04.

AZIMLI, A. (2020): "The impact of COVID-19 on the degree of dependence and structure of risk-return relationship: A quantile regression approach," *Finance Research Letters*, 36, 101648.

BÉLAND, L.-P., A. BRODEUR, D. MIKOLA, AND T. WRIGHT (2022): "The short-term economic consequences of COVID-19: occupation tasks and mental health in Canada," *Canadian Journal of Economics/Revue canadienne d'économique*, 55, 214–247.

BÉLAND, L.-P., A. BRODEUR, AND T. WRIGHT (2023): "The short-term economic consequences of Covid-19: exposure to disease, remote work and government response," *Plos one*, 18, e0270341.

BOCCALETTI, S., W. DITTO, G. MINDLIN, AND A. ATANGANA (2020): "Modeling and forecasting of epidemic spreading: The case of Covid-19 and beyond," *Chaos, solitons, and fractals*, 135, 109794.

BRODEUR, A., N. COOK, AND T. WRIGHT (2021): "On the effects of COVID-19 safer-at-home policies on social distancing, car crashes and pollution," *Journal of environmental economics and management*, 106, 102427.

BUHAI, S. (2005): "Quantile regression: overview and selected applications," *Ad Astra*, 4, 1–17.

CAR, Z., S. BARESSI ŠEGOTA, N. ANĐELIĆ, I. LORENCIN, AND V. MRZLJAK (2020): "Modeling the spread of COVID-19 infection using a multilayer perceptron," *Computational and mathematical methods in medicine*, 2020.

CHAKRABORTY, T. AND I. GHOSH (2020): "Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis," *Chaos, Solitons & Fractals*, 135, 109850.

CHAN, S., J. CHU, Y. ZHANG, AND S. NADARAJAH (2021): "Count regression models for COVID-19," *Physica A: Statistical Mechanics and its Applications*, 563, 125460.

CHEN, Y.-C. (2018): "Modal regression using kernel density estimation: A review," *Wiley Interdisciplinary Reviews: Computational Statistics*, 10, e1431.

CHEN, Y.-C., C. R. GENOVESE, R. J. TIBSHIRANI, AND L. WASSERMAN (2016): "Nonparametric modal regression," *The Annals of Statistics*, 44, 489–514.

CHO, J. S., T.-H. KIM, AND Y. SHIN (2015): "Quantile cointegration in the autoregressive distributed-lag modeling framework," *Journal of econometrics*, 188, 281–300.

DEMPSTER, A. P., N. M. LAIRD, AND D. B. RUBIN (1977): "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, 39, 1–22.

GNING, L., C. NDOUR, AND J. TCHUENCHE (2022): "Modeling COVID-19 daily cases in Senegal using a generalized Waring regression model," *Physica A: Statistical Mechanics and its Applications*, 597, 127245.

GUMAEI, A., M. AL-RAKHAMI, M. M. AL RAHHAL, F. ALBOGAMY, E. AL MAGHAYREH, AND H. AL-SALMAN (2021): "Prediction of COVID-19 confirmed cases using gradient boosting regression method," *Comput Mater Continua*, 66, 315–329.

GUPTA, S., L. MONTENOVO, T. NGUYEN, F. LOZANO-ROJAS, I. SCHMUTTE, K. SIMON, B. A. WEINBERG, AND C. WING (2023): "Effects of social distancing policy on labor market outcomes," *Contemporary Economic Policy*, 41, 166–193.

HAMERMESH, D. S. (2020): "Life satisfaction, loneliness and togetherness, with an application to Covid-19 lock-downs," *Review of Economics of the Household*, 18, 983–1000.

HE, G., Y. PAN, AND T. TANAKA (2020): "COVID-19, city lockdowns, and air pollution: evidence from China," *MedRxiv*.

KHARDANI, S. AND A. F. YAO (2017): "Non linear parametric mode regression," *Communications in Statistics-Theory and Methods*, 46, 3006–3024.

KRIEF, J. M. (2017): "Semi-linear mode regression," *The Econometrics Journal*, 20, 149–167.

LI, J., S. RAY, AND B. G. LINDSAY (2007): "A Nonparametric Statistical Approach to Clustering via Mode Identification," *Journal of Machine Learning Research*, 8, 1687–1723.

LU, H., P. NIE, AND L. QIAN (2021): "Do quarantine experiences and attitudes towards COVID-19 affect the distribution of mental health in China? A quantile regression analysis," *Applied Research in Quality of Life*, 16, 1925–1942.

MOLLALO, A., K. M. RIVERA, AND B. VAHEDI (2020): "Artificial neural network modeling of novel coronavirus (COVID-19) incidence rates across the continental United States," *International journal of environmental research and public health*, 17, 4204.

MUSULIN, J., S. BARESSI ŠEGOTA, D. ŠTIFANIĆ, I. LORENCIN, N. ANĐELIĆ, T. ŠUŠTERŠIČ, A. BLAGOJEVIĆ, N. FILIPOVIĆ, T. ĆABOV, AND E. MARKOVA-CAR (2021): "Application of artificial intelligence-based regression methods in the problem of covid-19 spread prediction: A systematic review," *International Journal of Environmental Research and Public Health*, 18, 4287.

OLMSTEAD, T. A. M. D. J. AND R. M. TERTILT (2020): "The impact of COVID-19 on gender equality," Tech. Rep. 26947, National Bureau of Economic Research.

PORTER, S. R. (2015): "Quantile regression: Analyzing changes in distributions instead of means," *Higher education: Handbook of theory and research*, 335–381.

ROJAS, F. L., X. JIANG, L. MONTENOVO, K. I. SIMON, B. A. WEINBERG, AND C. WING (2020): "Is the cure worse than the problem itself? Immediate labor market effects of COVID-19 case rates and school closures in the US," Tech. Rep. 27127, National Bureau of Economic Research.

SALGOTRA, R., M. GANDOMI, AND A. H. GANDOMI (2020a): "Evolutionary modelling of the COVID-19 pandemic in fifteen most affected countries," *Chaos, Solitons & Fractals*, 140, 110118.

——— (2020b): "Time series analysis and forecast of the COVID-19 pandemic in India using genetic programming," *Chaos, Solitons & Fractals*, 138, 109945.

SASAKI, H., T. SAKAI, AND T. KANAMORI (2020): "Robust modal regression with direct gradient approximation of modal regression risk," in *Proceedings of the Thirty-Sixth Conference on Uncertainty in Artificial Intelligence, UAI 2020, virtual online, August 3-6, 2020*, ed. by R. P. Adams and V. Gogate, AUAI Press, vol. 124 of *Proceedings of Machine Learning Research*, 380–389.

SCOTT, D. W. (1979): "On optimal and data-based histograms," *Biometrika*, 66, 605–610.

SHEATHER, S. J. AND M. C. JONES (1991): "A reliable data-based bandwidth selection method for kernel density estimation," *Journal of the Royal Statistical Society: Series B (Methodological)*, 53, 683–690.

SILVERMAN, B. W. (1986): *Density estimation for statistics and data analysis*, vol. 26, CRC press.

TUBADJI, A., F. BOY, AND D. WEBBER (2020): "Narrative economics, public policy and mental health," *Center for Economic Policy Research*, 20, 109–131.

ULLAH, A., T. WANG, AND W. YAO (2021): "Modal regression for fixed effects panel data," *Empirical Economics*, 60, 261–308.

——— (2022): "Nonlinear modal regression for dependent data with application for predicting COVID-19." *Journal of the Royal Statistical Society. Series A,(Statistics in Society)*, 185, 1424–1453.

VESPIGNANI, A., H. TIAN, C. DYE, J. O. LLOYD-SMITH, R. M. EGGO, M. SHRESTHA, S. V. SCARPINO, B. GUTIERREZ, M. U. KRAEMER, J. WU, ET AL. (2020): "Modelling covid-19," *Nature Reviews Physics*, 2, 279–281.

XIANG, S. AND W. YAO (2022): "Nonparametric statistical learning based on modal regression," *J. Comput. Appl. Math.*, 409, 114130.

YAO, W. AND L. LI (2014): "A new regression model: modal linear regression," *Scandinavian Journal of Statistics*, 41, 656–671.

YOUSEFPOUR, A., H. JAHANSHAHI, AND S. BEKIROS (2020): "Optimal policies for control of the novel coronavirus disease (COVID-19) outbreak," *Chaos, Solitons & Fractals*, 136, 109883.

YU, P., Z. ZHU, J. SHI, AND X. AI (2020): "Robust Estimation for Partial Functional Linear Regression Model Based on Modal Regression," *Journal of Systems Science and Complexity*, 33, 527–544.

ZHOU, H. AND X. HUANG (2016): "Nonparametric modal regression in the presence of measurement error," *Electronic Journal of Statistics*, 10, 3579–3620.

——— (2019): "Bandwidth selection for nonparametric modal regression," *Communications in Statistics-Simulation and Computation*, 48, 968–984.

ZIVKOVIC, M., N. BACANIN, A. DJORDJEVIC, M. ANTONIJEVIC, I. STRUMBERGER, T. A. RASHID, ET AL. (2021): "Hybrid genetic algorithm and machine learning method for covid-19 cases prediction," in *Proceedings of International Conference on Sustainable Expert Systems*, Springer, 169–184.

	Hospital Beds per 1,000 People (as of 2020)	Health Expenditure Growth between 2019 and 2020 (as % of GDP)	Cumulative Vaccinations per 100 People (as of Dec. 2022)	Fatality Rate (as of Dec,2022)
Canada	2.79	1.99% pt.	247.84	1.1% (48,349)
Japan	12.63	0.15% pt.	285.44	0.2% (50,827)
Korea	12.65	0.22% pt.	250.19	0.1% (30,975)
U.S.	2.80	2.14% pt.	197.37	1.1% (1,083,362)

Table 1: MEDICAL STATISTICS ON COVID-19 FOR THE FOUR COUNTRIES. The data on Hospital Beds is sourced from Trading Economics (https://tradingeconomics.com/country-list/hospital-beds). The data on Health Expenditure is sourced from The World Bank (https://data.worldbank.org/indicator). The data on cumulative vaccinations and fatality rate are sourced from Our World in Data (https://ourworldindata.org/covid-vaccinations).

	Method	Parameter $\backslash$ n	100	200	300	500	1,000
DGP1	MER	α_{1*}	0.3381	0.1625	0.1074	0.0661	0.0311
		β_{1*}	1.0970	0.5511	0.3604	0.2270	0.1035
	LQR	α_{2*}	0.3677	0.1735	0.1259	0.0717	0.0346
		β_{2*}	1.2485	0.6062	0.4221	0.2450	0.1170
	MR & SJ	α_{3*}	0.4563	0.3721	0.3721	0.1467	0.0923
		β_{3*}	0.9167	0.8285	0.7511	0.5574	0.2566
	MR & Scott	α_{3*}	0.4662	0.3792	0.3784	0.1462	0.0914
		β_{3*}	0.9257	0.8849	0.7663	0.5667	0.2673
	MR & Silverman	α_{3*}	0.5047	0.4511	0.4200	0.1531	0.1008
		β_{3*}	1.1022	1.0236	0.8921	0.6430	0.2592
DGP2	MER	α_{4*}	0.2541	0.1167	0.0791	0.0467	0.0231
		β_{4*}	0.7455	0.3613	0.2357	0.1331	0.0725
	LQR	α_{5*}	0.2818	0.1340	0.0918	0.0536	0.0271
		β_{5*}	0.8637	0.3867	0.2764	0.1669	0.0805
	MR & SJ	α_{6*}	1.9091	1.7029	1.3450	1.3166	1.2781
		β_{6*}	2.8691	2.5467	2.1244	2.0269	1.8912
	MR & Scott	α_{6*}	1.9723	1.7681	1.4101	1.4293	1.2901
		β_{6*}	2.7920	2.5719	2.2965	2.2910	1.9801
	MR & Silverman	α_{6*}	2.0121	1.7921	1.3771	1.3491	1.3291
		β_{6*}	2.8812	2.7021	2.3103	2.1310	1.9004

Table 2: THE MSES USING THE CROSS-SECTIONAL DATA SIMULATIONS. DGP1 is generated by simulating IID samples $\{(X_t, Y_t), i = 1, \ldots, n\}$ such that $Y_t = \alpha_* + \beta_* X_t + U_t$, where $\alpha_* = 0$, $\beta_* = 2$, and $U_t \sim 0.5N(-2, 3^2) + 0.5N(2, 1^2)$, from which $\mathbb{E}(Y_t \mid X_i) = 2X_t$, median $(Y_t \mid X_i) = 2X_t + 1$, and mode $(Y_t \mid X_i) = 2X_t + 2$. DGP2 is generated by simulating IID samples such that $Y_t = \alpha_* + \beta_* X_t + U_t$, where $\alpha_* = 0$, $\beta_* = 1$, and $U_t \sim \mathcal{X}_3^2 - 3$, from which $\mathbb{E}(Y_t \mid X_i) = X_t$, median $(Y_t \mid X_i) = X_t - 0.63$, and mode $(Y_t \mid X_i) = X_t - 2$. The simulation results are obtained by conducting 1,000 replications.

	Method	Parameter\n	100	200	300	500	1000
DGP3	MEAR	α_{7*}	0.0989	0.0391	0.0209	0.0046	0.0015
		β_{7*}	0.0008	0.0001	0.0000	0.0000	0.0000
	LQAR	α_{8*}	0.0844	0.0300	0.0154	0.0065	0.0005
		β_{8*}	0.0005	0.0001	0.0000	0.0000	0.0000
	MAR & SJ	α_{9*}	0.4282	0.2647	0.2121	0.1895	0.0810
		β_{9*}	0.0008	0.0001	0.0000	0.0000	0.0000
	MAR & Scott	α_{9*}	0.4833	0.2693	0.2177	0.1932	0.0821
		β_{9*}	0.0006	0.0001	0.0000	0.0000	0.0000
	MAR & Silverman	α_{9*}	0.6093	0.3722	0.3112	0.2069	0.1293
		β_{9*}	0.0008	0.0001	0.0000	0.0000	0.0000
DGP4	MEAR	α_{10*}	0.0001	0.0000	0.0000	0.0000	0.0000
		β_{10*}	0.0007	0.0001	0.0001	0.0000	0.0000
	LQAR	α_{11*}	0.0004	0.0001	0.0000	0.0000	0.0000
		β_{11*}	0.0002	0.0001	0.0001	0.0000	0.0000
	MAR & SJ	α_{12*}	0.2108	0.1123	0.1112	0.0925	0.0581
		β_{12*}	0.0004	0.0000	0.0000	0.0000	0.0000
	MAR & Scott	α_{12*}	0.2495	0.1164	0.1142	0.0919	0.0628
		β_{12*}	0.0014	0.0000	0.0000	0.0000	0.0000
	MAR & Silverman	α_{12*}	0.3030	0.1348	0.1339	0.1112	0.0725
		β_{12*}	0.0004	0.0000	0.0000	0.0000	0.0000

Table 3: THE MSEs USING THE TIME-SERIES DATA SIMULATIONS. This table compares the performances from different estimations: MEAR, LQAR, and MAR methods based upon the SJ's, Scott's, and Silverman's bandwidths. The first simulated dataset $\{Y_t : t = 1, \ldots, n\}$ is obtained from the following DGP3: $Y_t = \alpha_* + \beta_* Y_{t-1} + U_t$, where $\alpha_* = 0.3$, $\beta_* = 1$, and $U_t \sim \Gamma(3,1) - 3$. From this, $E(Y_t \mid Y_{t-1}) = 0.3 + Y_{t-1}$, median $(Y_t \mid Y_{t-1}) = Y_{t-1}$, and mode $(Y_t \mid Y_{t-1}) = -0.7 + Y_{t-1}$. The second simulated dataset $\{Y_t : t = 1, \ldots, n\}$ is obtained from the following DGP4: $Y_t = \alpha_* + \beta_* Y_{t-1} + U_t$, where $\alpha_* = 0.3$, $\beta_* = 0.5$, and $U_t \sim \Gamma(3,1) - 3$. From this, $E(Y_t \mid Y_{t-1}) = 0.3 + 0.5Y_{t-1}$, median $(Y_t \mid Y_{t-1}) = 0.5Y_{t-1}$, and mode $(Y_t \mid Y_{t-1}) = -0.7 + 0.5Y_{t-1}$.

Country	Method	Training Set			Test Set		
		RMSE	MAE	CQF	RMSE	MAE	CQF
Canada	MER	0.0148	0.0118	0.0103	0.0318	0.0304	0.0298
	LQR	0.0152	0.0111	0.0085	0.0354	0.0342	0.0332
	MR	0.0148	0.0117	0.0102	0.0319	0.0306	0.0299
	BMR	0.0063	0.0051	**0.0038**	0.0709	0.0531	0.0466
	LPMR	0.0063	0.0052	0.0043	0.0171	0.0146	**0.0124**
Japan	MER	0.0303	0.0267	0.0275	0.0756	0.0749	0.0755
	LQR	0.0326	0.0250	0.0168	0.0838	0.0832	0.0838
	MR	0.0303	0.0265	0.0165	0.0758	0.0750	0.0757
	BMR	0.0033	0.0027	0.0024	0.0083	0.0069	**0.0045**
	LPMR	0.0031	0.0025	**0.0023**	0.0125	0.0115	0.0103
Korea	MER	0.0621	0.0551	0.0583	0.0961	0.0960	0.0958
	LQR	0.0644	0.0543	0.0537	0.0643	0.0639	0.0619
	MR	0.0706	0.0555	0.0496	0.0550	0.0541	0.0517
	BMR	0.0043	0.0028	**0.0014**	0.0823	0.0690	0.0532
	LPMR	0.0036	0.0026	0.0018	0.0870	0.0605	**0.0399**
U.S.	MER	0.0646	0.0531	0.0490	0.1238	0.1221	0.1295
	LQR	0.0722	0.0486	0.0332	0.0912	0.0898	0.0947
	MR	0.0749	0.0489	0.0336	0.0779	0.0765	0.0805
	BMR	0.0072	0.0049	0.0035	0.0283	0.0221	**0.0147**
	LPMR	0.0070	0.0047	**0.0030**	0.0251	0.0204	**0.0147**

Table 4: THE QUANTITATIVE RESULTS OF DIFFERENT ESTIMATION METHODS. This table evaluates the efficacy of different regression methods for forecasting the cumulative confirmed COVID-19 cases from February 8, 2022, to April 8, 2022. The results are measured by RMSE, MAE, and CQF.

Country	Before			After		
	Dickey-Fuller	p-value	Reject	Dickey-Fuller	p-value	Reject
Canada	-3.0538	0.1516	No	-6.4527	< 0.01	Yes
Japan	-0.9009	0.9447	No	-5.4706	< 0.01	Yes
Korea	-1.9200	0.6064	No	-5.5924	< 0.01	Yes
U.S.	-4.0388	0.0151	No	-5.8077	< 0.01	Yes

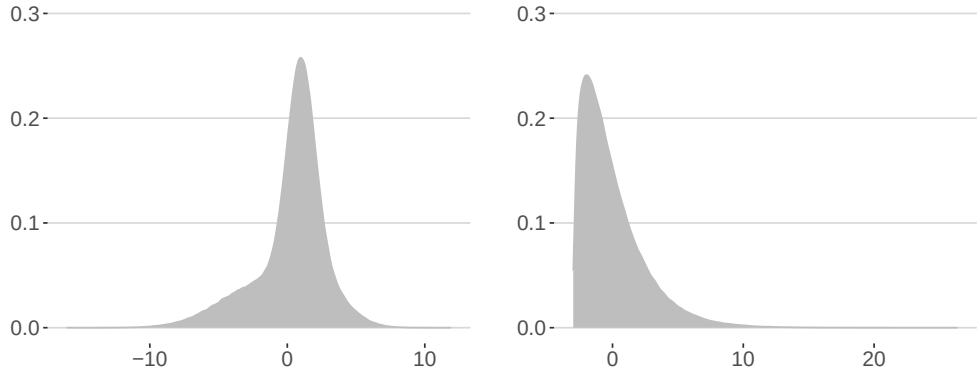
Table 5: THE ADF TEST RESULTS BEFORE AND AFTER DIFFERENCING THE ACCUMULATED CONFIRMED COVID-19 CASES. The considered dataset ranges from February 8, 2022, to April 8, 2022, and the ADF test significance level is set to 0.01. To ensure the stationarity of the time series, we set the difference order as 3.

Method			Canada	Japan	Korea	U.S.
SJ	RMSE	Training Set	**0.1764**	**0.1350**	**0.1085**	0.0355
		Test Set	**0.2857**	**0.0699**	**0.0701**	**0.0307**
	Coef	α_*	-0.0236	-0.0044	0.0000	0.0029
		β_{1*}	-0.5224	-0.9021	-0.8679	-0.5925
		β_{2*}	-0.3358	-0.8722	-1.0773	-0.6856
		β_{3*}	-0.6016	-0.8328	-1.3416	-0.7422
		β_{4*}	-1.1054	-0.7654	-1.3734	-0.8425
		β_{5*}	-1.3693	-0.8430	-1.4871	-0.9307
		β_{6*}	-1.3588	-0.7832	-1.2355	-0.9529
Scott	RMSE	Training Set	0.1779	0.1450	0.1279	0.0355
		Test Set	0.2889	0.1044	0.0929	0.0308
	Coef	α_*	-0.0234	0.0526	0.0035	0.0029
		β_{1*}	-0.5289	-0.4999	-1.1798	-0.5953
		β_{2*}	-0.3329	-0.9829	-1.0942	-0.6893
		β_{3*}	-0.6020	-1.0668	-1.2720	-0.7473
		β_{4*}	-1.1118	-1.1104	-1.1384	-0.8464
		β_{5*}	-1.3744	-1.0587	-1.3406	-0.9316
		β_{6*}	-1.3638	-0.9479	-1.0740	-0.9527
Silverman	RMSE	Training Set	0.1799	0.1506	0.1093	**0.0354**
		Test Set	0.2907	0.1069	0.0806	0.0308
	Coef	α_*	-0.0235	0.0509	-0.0088	0.0025
		β_{1*}	-0.5364	-0.4543	-0.5419	-0.5885
		β_{2*}	-0.3288	-0.9954	-0.6828	-0.6800
		β_{3*}	-0.5997	-1.1023	-0.9043	-0.7340
		β_{4*}	-1.1199	-1.1617	-0.7628	-0.8361
		β_{5*}	-1.3785	-1.0871	-1.1533	-0.9277
		β_{6*}	-1.3723	-0.9580	-0.9320	-0.9529

Table 6: THE QUANTITATIVE RESULTS OF DIFFERENT BANDWIDTH SELECTION METHODS FOR COVID-19. The RMSE is computed by comparing the forecasts with the realized confirmed cases. Here, α_* denotes the intercept; and the lag coefficients are denoted as β_{1*}, β_{2*}, β_{3*}, β_{4*}, β_{5*}, and β_{6*}.

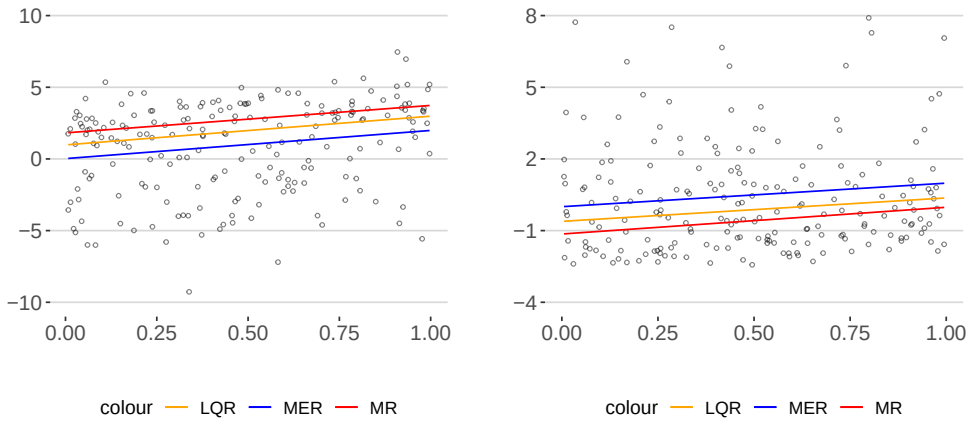
Country	Method	Training Set			Test Set (Teacher Forcing)			Test Set (Non-teacher Forcing)		
		RMSE	MAE	CQF	RMSE	MAE	CQF	RMSE	MAE	CQF
Canada	MEAR	0.1443	0.1145	0.1043	0.2206	0.1968	**0.1784**	0.2745	0.2243	0.1991
	LQAR	0.1487	0.1120	0.0930	0.2494	0.2124	0.1898	0.2373	0.2004	0.1893
	MAR	0.1764	0.1256	**0.0664**	0.2857	0.2338	0.1855	0.1981	0.1794	**0.1855**
Japan	MEAR	0.1151	0.0906	0.0656	0.0625	0.0535	0.0492	0.2359	0.2185	0.2776
	LQAR	0.1201	0.0834	0.0510	0.0609	0.0487	0.0416	0.3254	0.3039	0.2822
	MAR	0.1350	0.0883	**0.0404**	0.0699	0.0538	**0.0396**	0.3796	0.3574	**0.2541**
Korea	MEAR	0.1027	0.0647	0.0348	0.0757	0.0593	0.0400	0.0675	0.0532	0.0748
	LQAR	0.1059	0.0624	0.0349	0.0967	0.0791	0.0806	0.0835	0.0707	0.0819
	MAR	0.1085	0.0654	**0.0217**	0.0701	0.0506	**0.0360**	0.0781	0.0653	**0.0654**
U.S.	MEAR	0.0342	0.0256	0.0197	0.0303	0.0211	0.0136	0.0302	0.0225	**0.0135**
	LQAR	0.0348	0.0247	0.0189	0.0315	0.0219	0.0168	0.0463	0.0295	0.0180
	MAR	0.0355	0.0251	**0.0157**	0.0307	0.0198	**0.0133**	0.0291	0.0212	0.0141

Table 7: THE QUANTITATIVE RESULTS OF THE FORECAST METHODS. This table compares the performances of the MEAR, LQAR, and MAR methods applied to the cumulative confirmed COVID-19 cases. The MAR model utilizes SJ's bandwidth. The data range from February 8, 2022, to April 8, 2022. For the test dataset, the teacher-forcing and non-teacher-forcing methods are separately applied.



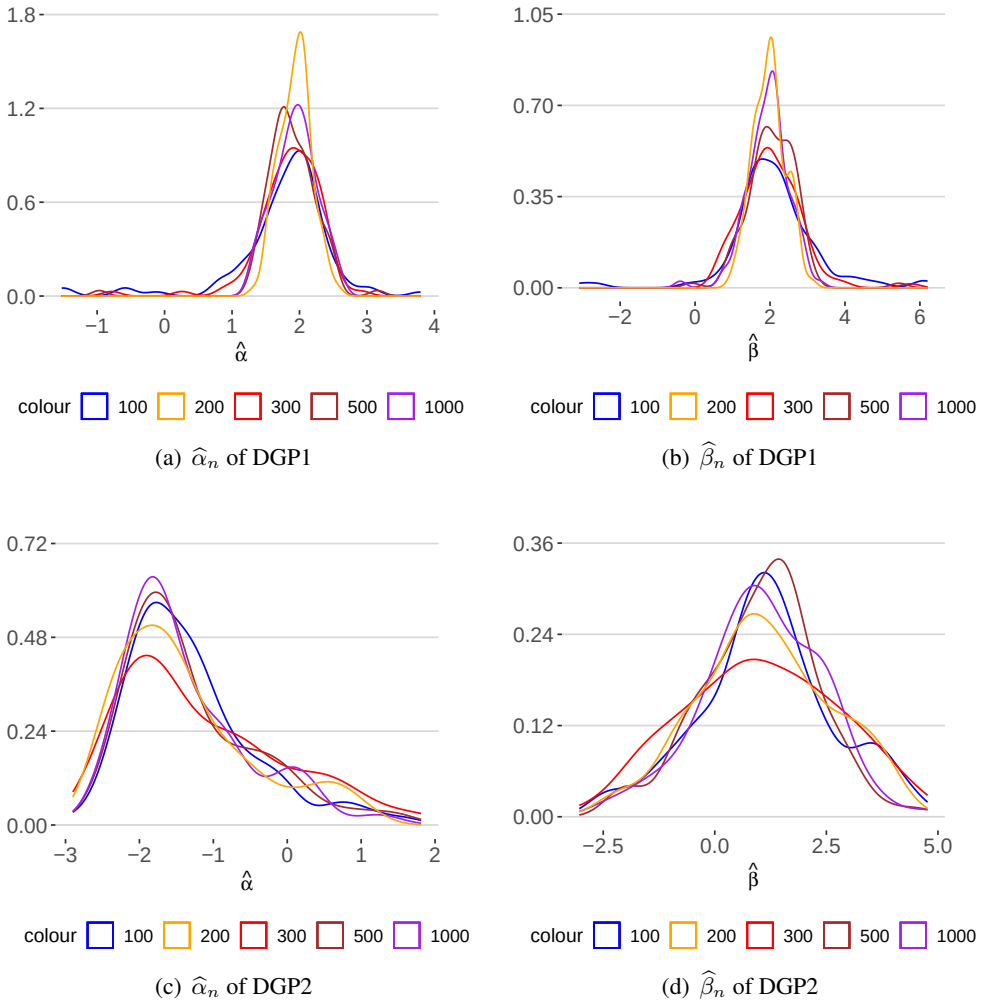
(a) DGP1 (b) DGP2

Figure 1: THE PROBABILITY DENSITY FUNCTION OF THE ERROR DISTRIBUTION IN THE DGPS. The skewed error of DGP1 is generated from $U_t \sim 0.5N(-2, 3^2) + 0.5N(2, 1^2)$. The skewed error of DGP2 is generated from $U_t \sim \mathcal{X}_3^2 - 3$.



(a) DGP1 (b) DGP2

Figure 2: THE SCATTER PLOT AND REGRESSION LINES. 200 data points are generated from DGP1: $Y_t = \alpha_* + \beta_* X_t + U_t$, where $\alpha_* = 0$, $\beta_* = 2$ and $U_t \sim 0.5N(-2, 3^2) + 0.5N(2, 1^2)$. The red, orange, and blue lines represent the conditional mode, median, and mean functions, respectively. Other 200 data points are generated from DGP2: $Y_t = \alpha_* + \beta_* X_t + U_t$, where $\alpha_* = 0$, $\beta_* = 1$ and $U_t \sim \mathcal{X}_3^2 - 3$. The red, orange, and blue lines represent the modal, median, and mean regressions, respectively.



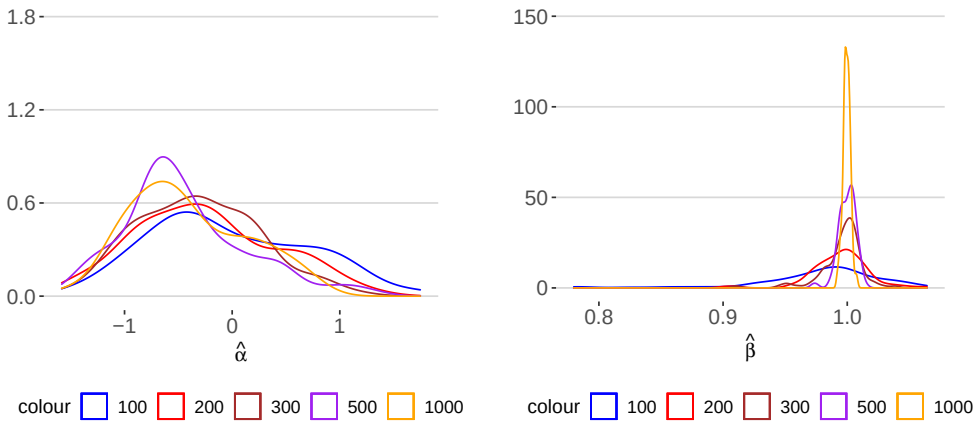
(a) $\widehat{\alpha}_n$ of DGP1

(b) $\widehat{\beta}_n$ of DGP1

(c) $\widehat{\alpha}_n$ of DGP2

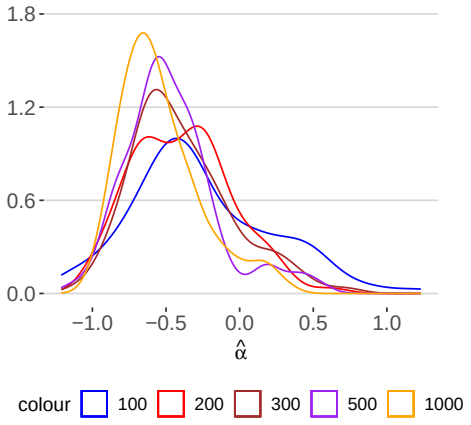
(d) $\widehat{\beta}_n$ of DGP2

Figure 3: EMPIRICAL DENSITY OF THE MODAL REGRESSION COEFFICIENTS FROM THE CROSS-SECTIONAL DATA. The four figures separately display the probability densities of the estimated coefficients $\widehat{\alpha}_n$ and $\widehat{\beta}_n$ using the cross-sectional data described as DGP 1 and DGP2.

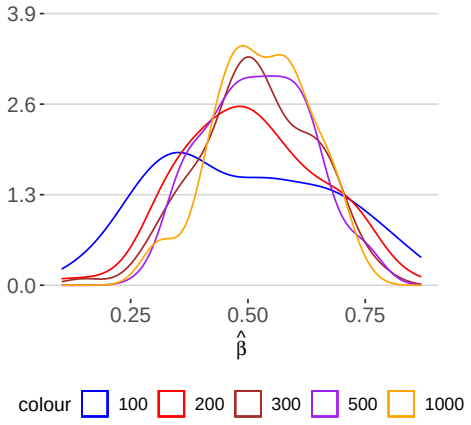


(a) $\widehat{\alpha}_n$ of DGP3

(b) $\widehat{\beta}_n$ of DGP3



(c) $\widehat{\alpha}_n$ of DGP4



(d) $\widehat{\beta}_n$ of DGP4

Figure 4: EMPIRICAL DENSITY OF THE MODAL REGRESSION COEFFICIENTS FROM THE TIME-SERIES DATA. The four figures separately display the probability densities of the estimated coefficients $\widehat{\alpha}_n$ and $\widehat{\beta}_n$ using the time-series data described as DGP 3 and DGP 4.

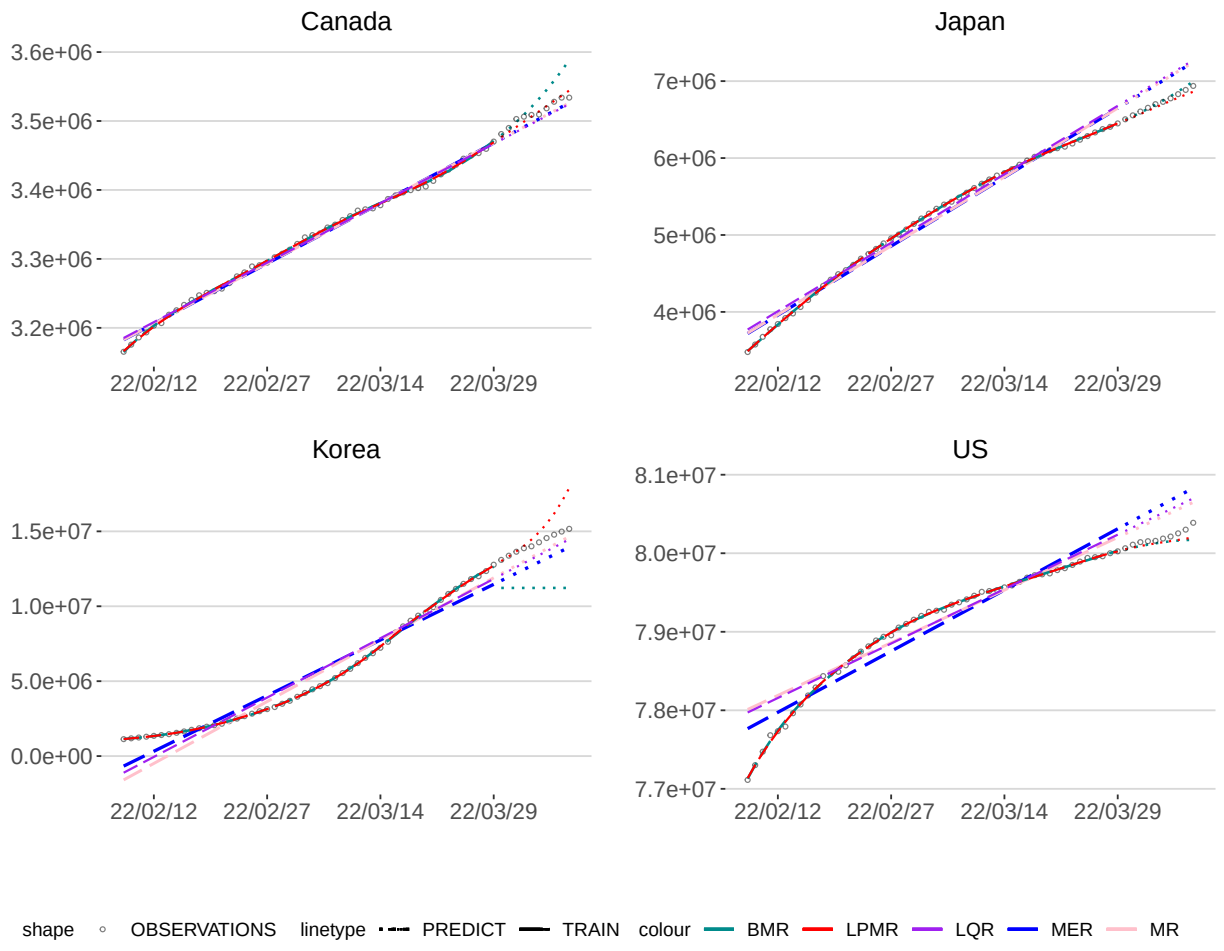
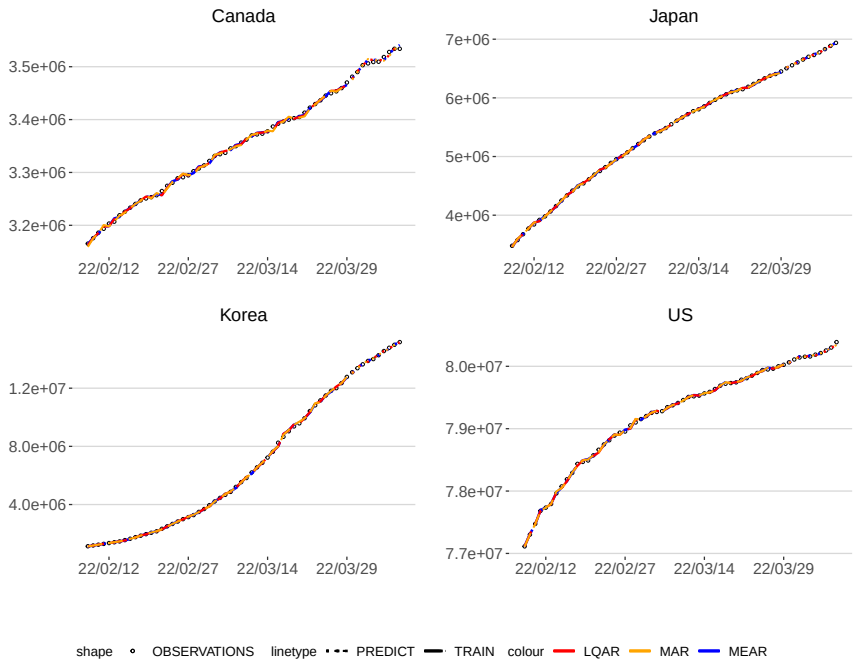
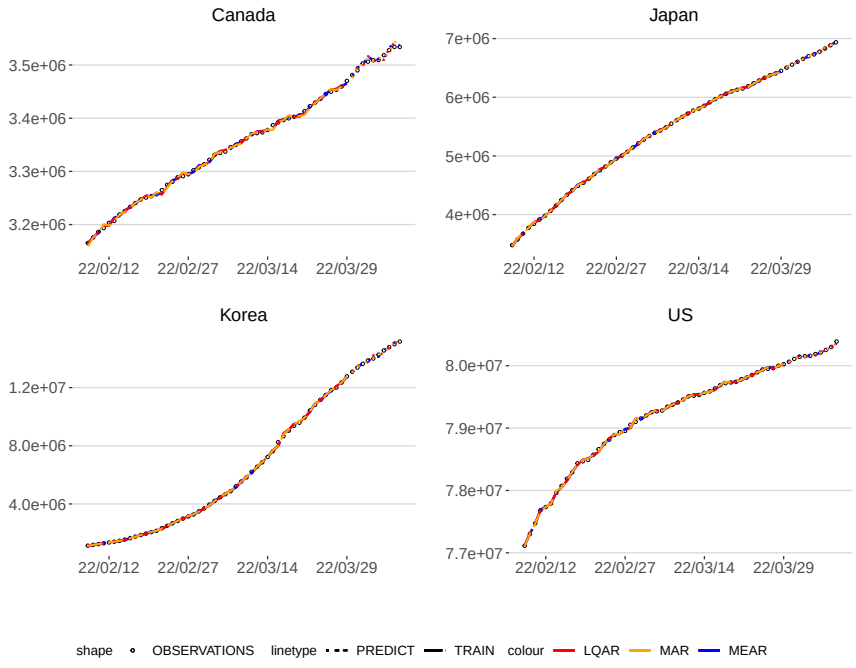


Figure 5: THE SCATTER PLOT OF ALL DATA POINTS AND THE REGRESSION LINES FROM THE DIFFERENT REGRESSION METHODS. This figure shows different forecasts from the five regression methods: MER, LQR, MR, BMR, and LPMR. The vertical axis denotes the cumulative confirmed COVID-19 cases. The real observations, in-sample and out-of-sample forecasts are denoted by circles, 'PREDICT' and 'TRAIN', respectively. Different colors are used to indicate different regression methods.

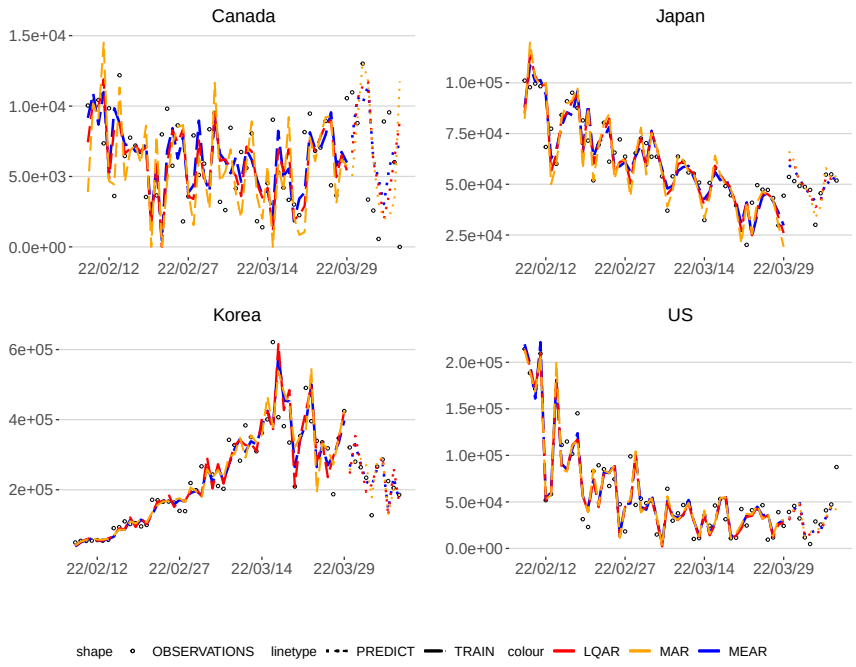


(a) Teacher-Forcing Method

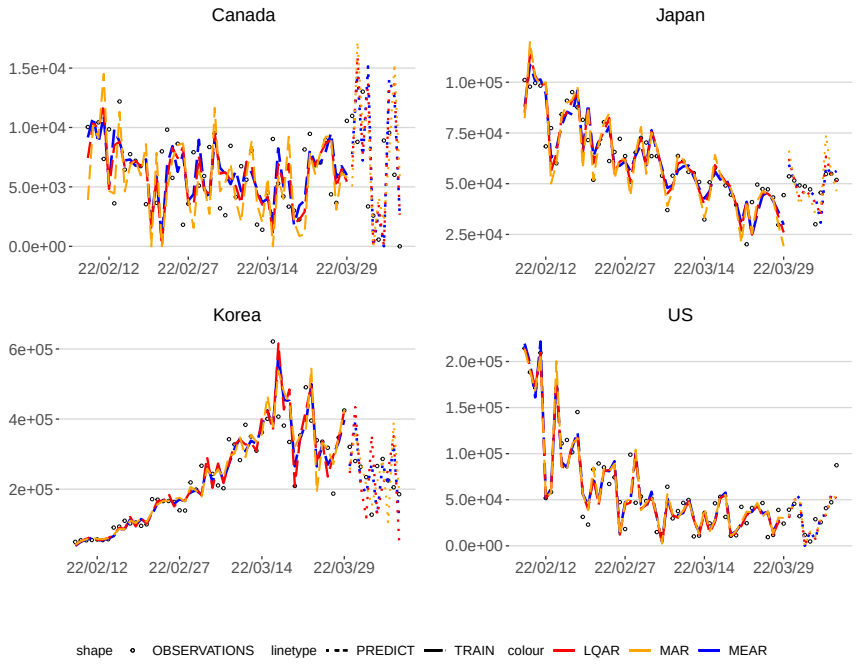


(b) Non-Teacher-Forcing Method

Figure 6: THE QUALITATIVE FORECASTING RESULTS OF THE MEAR, LQAR, AND MAR METHODS ON THE ACCUMULATED CONFIRMED CASES. The real observations, in-sample and out-of-sample forecasts are denoted by circles, 'PREDICT' and 'TRAIN', respectively. Different colors are used to indicate different methods. The red, orange, and blue lines represent the forecasting made by the MEAR, LQAR, and MAR methods, respectively.

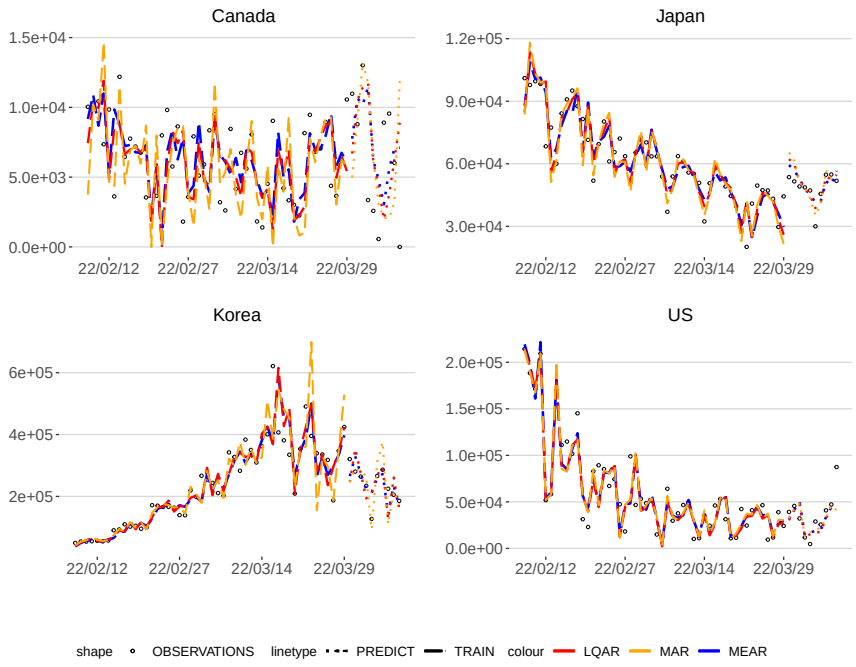


(a) Teacher-Forcing Method

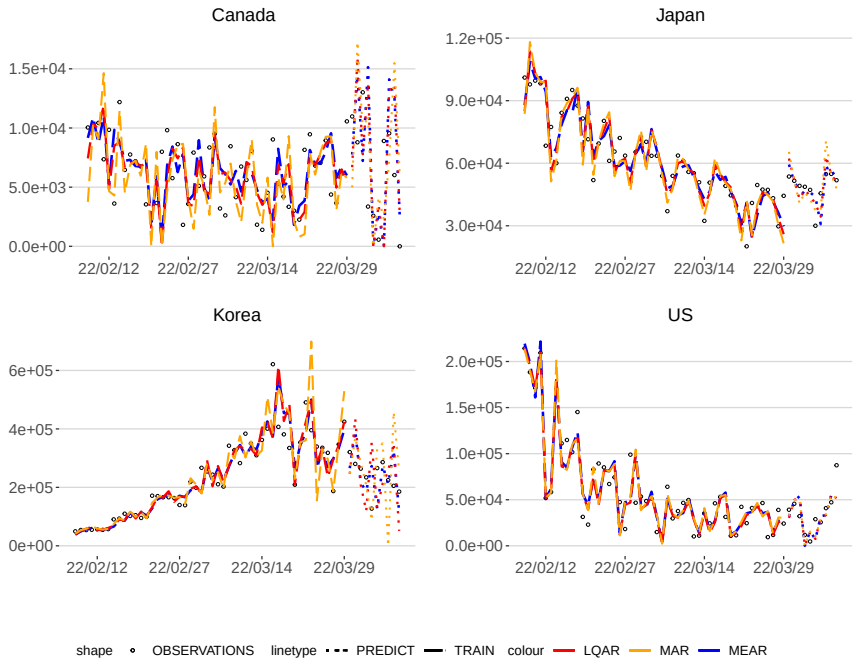


(b) Non-Teacher-Forcing Method

Figure 7: THE QUALITATIVE FORECASTING RESULTS OF THE MEAR, LQAR, AND MAR METHODS ON THE DAILY CONFIRMED CASES. The real observations, in-sample and out-of-sample forecasts are denoted by circles, 'PREDICT' and 'TRAIN', respectively. Different colors are used to indicate different methods. The blue, red, and orange lines represent the forecasting made by the MEAR, LQAR, and MAR methods, respectively. SJ's bandwidth is used.

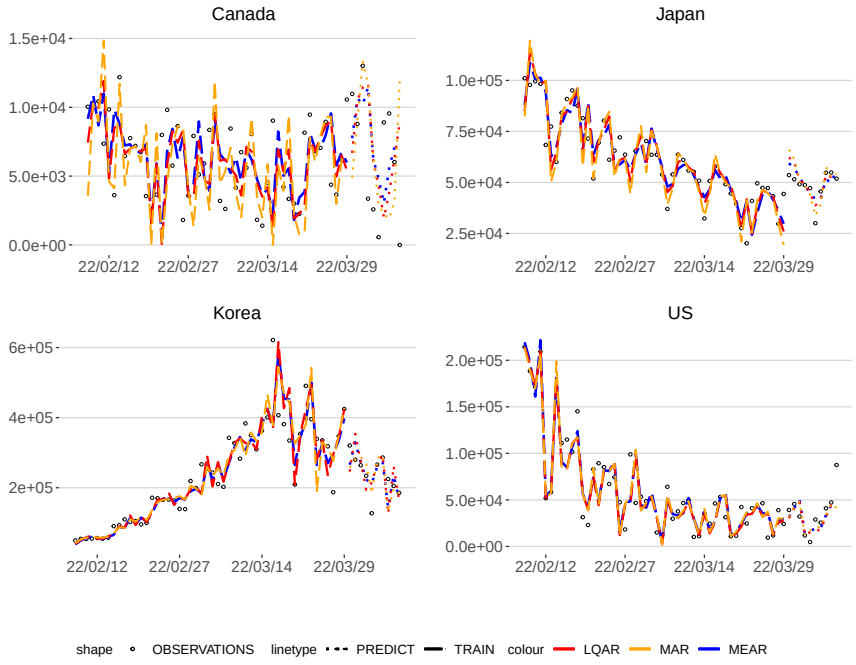


(a) Teacher-Forcing Method

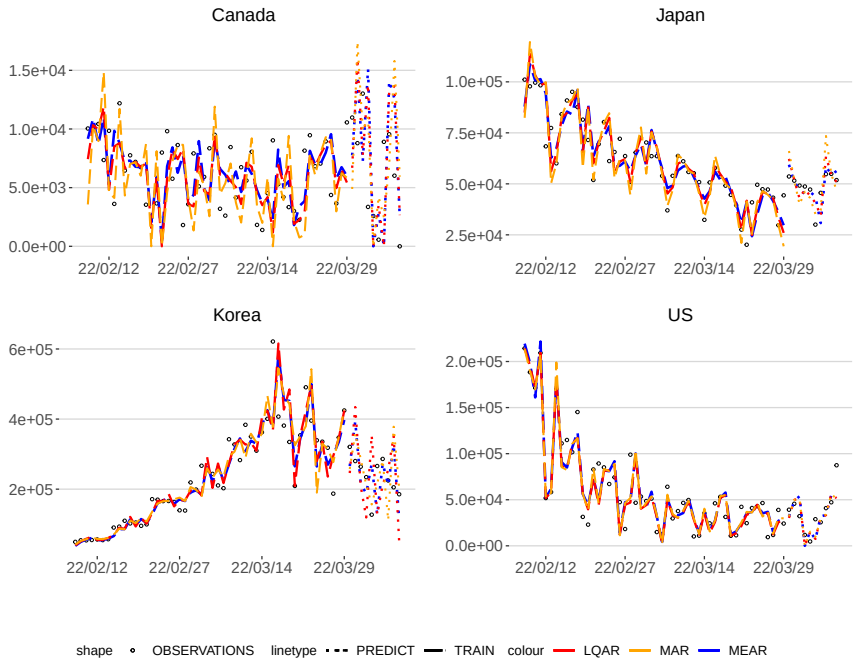


(b) Non-Teacher-Forcing Method

Figure 8: THE QUALITATIVE FORECASTING RESULTS OF THE MEAR, LQAR, AND MAR METHODS ON THE DAILY CONFIRMED CASES. The real observations, in-sample and out-of-sample forecasts are denoted by circles, 'PREDICT' and 'TRAIN', respectively. Different colors are used to indicate different methods. The blue, red, and orange lines represent the forecasting made by the MEAR, LQAR, and MAR methods, respectively. Scott's bandwidth is used.



(a) Teacher-Forcing Method



(b) Non-Teacher-Forcing Method

Figure 9: THE QUALITATIVE FORECASTING RESULTS OF THE MEAR, LQAR, AND MAR METHODS ON THE DAILY CONFIRMED CASES. The real observations, in-sample and out-of-sample forecasts are denoted by circles, 'PREDICT' and 'TRAIN', respectively. Different colors are used to indicate different methods. The blue, red, and orange lines represent the forecasting made by the MEAR, LQAR, and MAR methods, respectively. Silverman's bandwidth is used.

Online Supplement for 'Forecasting the Confirmed COVID-19 Cases Using Modal Regression'*

by

Xin Jing[a] and Jin Seo Cho[a]

[a]Yonsei University

In the current Appendix, we prove that $Q_{n,h}(\theta^{(k+1)}) \geq Q_{n,h}(\theta^{(k)})$ and provide the descriptive statistics of the confirmed COVID-19 cases.

A Appendix

Proof of $Q_{n,h}(\theta^{(k+1)}) \geq Q_{n,h}(\theta^{(k)})$: For each iteration of the MEM algorithm described in Section 3.1, the optimized objective function outcome gradually increases. In other words, for any positive integer k, $Q_{n,h}(\theta^{(k+1)}) \geq Q_{n,h}(\theta^{(k)})$. The proof is as follows: we note that

$$
\begin{aligned}
&\log Q_{n,h}\left(\theta^{(k+1)}\right) - \log Q_{n,h}\left(\theta^{(k)}\right) \\
&= \log \sum_{t=1}^{n} \phi_h\left(Y_t - m(X_t, \theta^{(k+1)})\right) - \log \sum_{t=1}^{n} \phi_h\left(Y_t - m(X_t, \theta^{(k)})\right) \\
&= \log \left[\sum_{t=1}^{n} \frac{\phi_h\left(Y_t - m(X_t, \theta^{(k+1)})\right)}{\sum_{j=1}^{n} \phi_h\left(Y_j - m(X_j, \theta^{(k)})\right)}\right] \\
&= \log \left[\sum_{t=1}^{n} \frac{\phi_h\left(Y_t - m(X_t, \theta^{(k)})\right)}{\sum_{t=1}^{n} \phi_h\left(Y_t - m(X_t, \theta^{(k)})\right)} \frac{\phi_h\left(Y_t - m(X_t, \theta^{(k+1)})\right)}{\phi_h(Y_t - m(X_t, \theta^{(k)}))}\right] \\
&= \log \left[\sum_{t=1}^{n} \pi(t \mid \theta^{(k)}) \frac{\phi_h\left(Y_t - m(X_t, \theta^{(k+1)})\right)}{\phi_h\left(Y_t - m(X_t, \theta^{(k)})\right)}\right]
\end{aligned}
$$

by noting that

$$
\pi(t \mid \theta^{(k)}) = \frac{\phi_h\left(Y_t - m(X_t, \theta^{(k)})\right)}{\sum_{t=1}^{n} \phi_h\left(Y_t - m(X_t, \theta^{(k)})\right)}.
$$

*The editor-in-chief, Derek Bunn, and an anonymous referee provided very helpful comments for which we are most grateful. The authors are grateful to Jihye Jung, Jaeseung Lee, and Moo Hyun Yang for their constructive comments. Jing and Cho acknowledge research support from the Yonsei University Research Grant of 2023.

From this, we obtain

$$\log Q_h\left(\theta^{(k+1)}\right) - \log Q_h\left(\theta^{(k)}\right) \geq \sum_{t=1}^{n} \pi(t \mid \theta^{(k)}) \log\left\{\frac{\phi_h\left(Y_t - m(X_t, \theta^{(k+1)})\right)}{\phi_h\left(Y_t - m(X_t, \theta^{(k)})\right)}\right\}$$

by applying Jensen's inequality. If we further apply the definition of $\theta^{(k+1)}$ from the M-step,

$$\sum_{t=1}^{n} \pi(t \mid \theta^{(k)}) \log\left\{\phi_h\left(Y_t - m(X_t, \theta^{(k+1)}\right))\right\} \geq \sum_{t=1}^{n} \pi(t \mid \theta^{(k)}) \log\left\{\phi_h\left(Y_t - m(X_t, \theta^{(k)}\right))\right\},$$

implying that

$$\log\left\{Q_h\left(\theta^{(k+1)}\right)\right\} - \log\left\{Q_h\left(\theta^{(k)}\right)\right\} \geq 0.$$

This completes the proof. □

We next provide Table A.1 containing the descriptive statistics of the three-times differenced confirmed COVID-19 cases for Canada, Japan, South Korea, and the U.S. The confirmed cases are transformed according to the min-max normalization.

Country	Mean	Median	Max	Min	Std	Skewness	Kurtosis	n
Canada	-0.0173	0.0483	0.9223	-1.0378	0.3863	-0.31323	-0.1602	60
Japan	0.0040	-0.0476	0.6807	-0.4512	0.2319	0.9643	0.5565	60
Korea	-0.0006	-0.0049	0.4967	-0.7421	0.1947	-0.3592	2.9991	60
U.S.	-0.0116	0.0021	0.4792	-1.1445	0.2382	-1.9729	7.1857	60

Table A.1: DESCRIPTIVE STATISTICS OF THE THREE-TIMES DIFFERENCED CONFIRMED COVID-19 CASES FOR CANADA, JAPAN, SOUTH KOREA, AND THE U.S. Descriptive statistics are computed over 60 days from February 8, 2022, to April 8, 2022. The data are available at the following URL: https://github.com/CSSEGISandData/COVID-19.