JUSTC

Article

http://justc.ustc.edu.cn

Received: February 12, 2025; Accepted: March 30, 2025

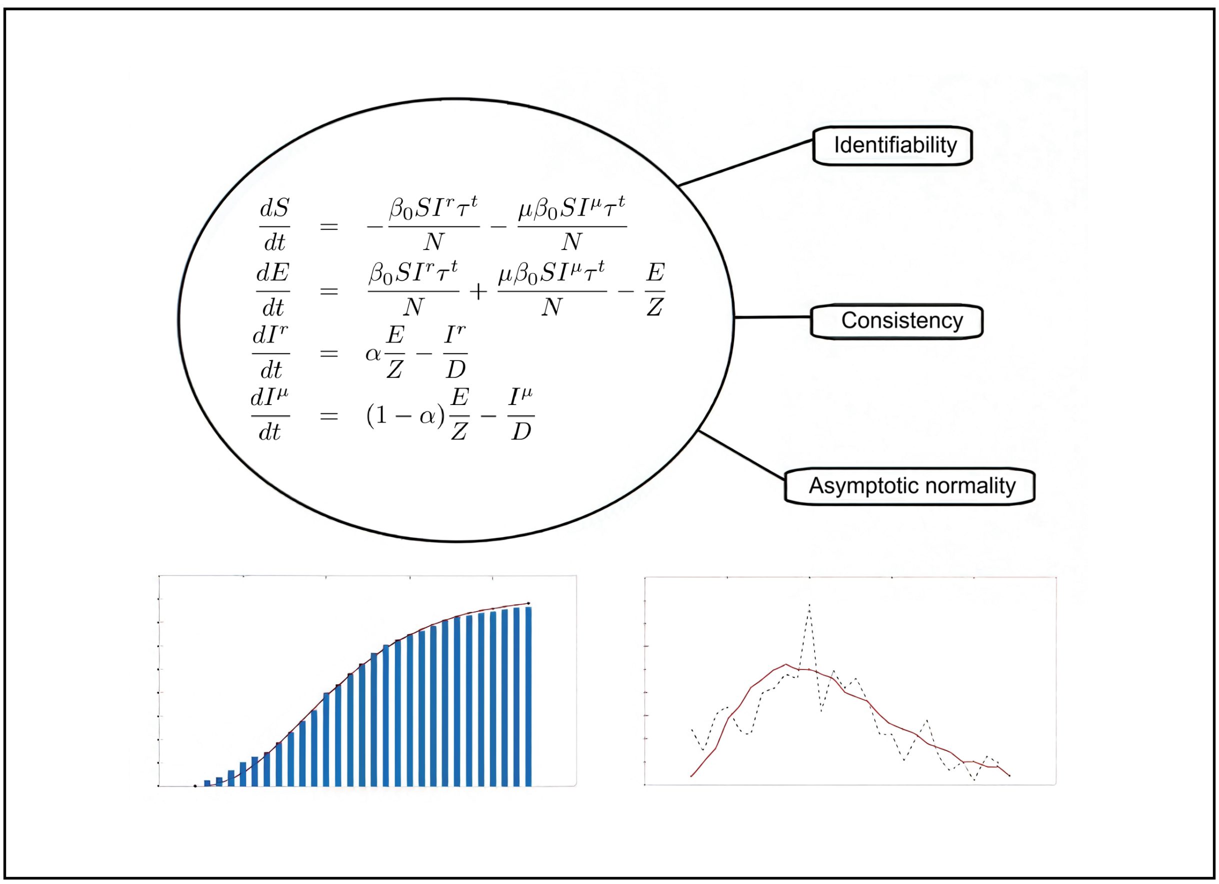
Theoretical foundations and application of the varying coefficient SEIR model

Tianyi Sun, and Baisuo Jin ✉

School of Management, University of Science and Technology of China, Hefei 230026, China

✉Correspondence: Baisuo Jin, E-mail: jbs@ustc.edu.cn

Graphical abstract



The theoretical properties of the varying coefficient SEIR model including identifiability and asymptotic properties under discussion.

Public summary

- This study provides a rigorous theoretical foundation for an attenuation-rate-based varying coefficient SEIR model, enabling systematic assessment of epidemic suppression under government interventions.

- Theoretical analysis addressed three fundamental challenges: parameter identifiability, consistency of estimators, and asymptotic normality, providing statistical guarantees for reliable epidemic inference.

- Empirical validation using COVID-19 data of 2020 demonstrated the model's practical applicability, with simulated daily case trajectories closely matching real-world observations.

Citation: Sun T Y, Jin B S. Theoretical foundations and application of the varying coefficient SEIR model. *JUSTC*, **2025**, 55(0): . CSTR: 32290.14.JUSTC-2025-0014. DOI: 10.52396/JUSTC-2025-0014

JUSTC

Article

http://justc.ustc.edu.cn

Received: February 12, 2025; Accepted: March 30, 2025

Theoretical foundations and application of the varying coefficient SEIR model

Tianyi Sun, and Baisuo Jin ✉

School of Management, University of Science and Technology of China, Hefei 230026, China

✉Correspondence: Baisuo Jin, E-mail: jbs@ustc.edu.cn

Cite This: *JUSTC*, **2025**, 55(X): (7pp)   Read Online   Supporting Information

Abstract: In recent years, various infectious diseases, including coronavirus disease 2019 (COVID-19), have occurred many times around the world, which makes the statistical modeling and analysis of the epidemic particularly important. The modeling and inference of epidemics under government control measures are still worthy of research in this field. In this work, we establish the theoretical foundation for the varying coefficient SEIR model with an attenuation rate, which is built to assess precisely the attenuation stages of the epidemic under control measures. For the model, we discuss the identifiability of the model parameters, analyze the consistency and asymptotic normality of the parameter estimates, and provide the corresponding proofs. Finally, we apply the model to COVID-19 outbreaks in 2020 and simulate daily new cases with the inferred parameters. The simulated case number curve is close to the real data, indicating the usability of the model.

Keywords: global epidemic; varying coefficient SEIR model; identifiability; consistency; asymptotic normality

CLC number: O213.9   **Document code:** A

2020 Mathematics Subject Classification: 62P99

1 Introduction

Epidemic models have long been a central focus in the study of dynamic system modeling. The classical mathematical frameworks, particularly the susceptible-infected-recovered (SIR) model[1] and susceptible-exposed-infected-recovered (SEIR) model[2], along with their various extensions[3–8], have been widely applied to characterize the transmission dynamics of infectious diseases. The emergence of coronavirus disease 2019 (COVID-19) significantly accelerated the development and analysis of epidemic models globally. Since 2020, the virus and its variants have spread widely around the world, leading to a large number of confirmed cases[9]. The high transmission rate and evolving nature of the virus pose substantial challenges for disease control and patient care[10, 11].

Given the profound social and economic impacts of the COVID-19 pandemic, extensive research has been conducted across medical and nursing domains. Statistical modeling approaches have focused particularly on estimating key epidemiological parameters, including the transmission rate, recovery duration, and other characteristics across different regions[12–18]. To analyze the trend of epidemic development over time[19, 20], statistical models, as well as classical epidemic models, have been employed at various stages of the outbreak.

With the spread of the epidemic around the world, governments of various countries and regions have implemented targeted control measures to varying degrees, including restricting transportation, closing schools and businesses, and, in some cases, locking down entire cities, such measures aimed at curbing the transmission of the virus. In addition to governmental actions, public health initiatives emphasize individual behavioral modifications, particularly through reduced social interactions and widespread adoption of face mask usage. Empirical evidence suggests that these containment strategies significantly attenuated the impact of the pandemic. Building proper models to quantify the causal effects of containment measures on COVID-19 transmission dynamics is valuable. In 2024, Ref. [21] proposed a novel SEIR model incorporating time-varying coefficients, which quantifies the effectiveness of government control measures by considering the attenuation rate of disease transmission. The purpose of this model is to capture the evolving epidemic trajectory following the implementation of containment measures and to evaluate the effects of the control measures in different cities. The effectiveness of both the model and the algorithm was validated through simulation studies and an analysis of real data spanning 136 Chinese cities.

While the original work demonstrated the model's empirical performance, it remained silent on the underlying theoretical foundations. Therefore, in this paper, we establish the theoretical basis for the varying coefficient SEIR (VCSEIR) model, focusing on the identifiability of the model, the consistency of parameter estimation, and the asymptotic normality of the parameter estimates. To complement these theoretical advancements, we further demonstrate the model's robustness through extended applications to broader empirical datasets, showcasing its fitting capabilities across various epidemic scenarios.

The remainder of this paper is organized as follows. In Section 2, we provide a detailed introduction to the VCSEIR model and the iterative algorithm used for estimating the model parameters. In Section 3, we present the theorems that

CSTR: 32290.14.JUSTC-2025-0014

DOI: 10.52396/JUSTC-2025-0014

JUSTC, **2025**, 55(X):

establish the theoretical properties of the model. Section 4 demonstrates the application of the model and algorithm to the 2020 COVID-19 outbreak in various countries and regions. The proofs of the theorems are provided in the Supporting Information.

2 Model

In this section, we introduce the VCSEIR model proposed by Ref. [21]. The model was built on the basis of the work of Ref. [3], with a new parameter defined as the attenuation rate over time, the model aims to infer the attenuation stage of the epidemic.

The original SEIR model[2] employs a four-compartment structure: susceptible, exposed, infected, and recovered compartments to simulate disease progression through infection, incubation, and recovery phases. However, this conventional framework has limitations in modeling COVID-19 transmission dynamics, particularly given the heightened transmissibility of the virus and the substantial proportion of mild and asymptomatic cases that are difficult to diagnose. Therefore, to address these empirical challenges, the VCSEIR model replaces the traditional infected compartment with distinct confirmed and unconfirmed case categories. Unconfirmed cases have milder symptoms and are therefore not recorded in the official surveillance systems, and the model reasonably assumes that unconfirmed cases have lower infectivity weights than confirmed cases do.

Within the framework of epidemic modeling, the transmission rate constitutes a fundamental parameter, representing the average number of individuals with whom a single infected person can transmit the disease within a given time period, typically one day. This parameter serves as an indicator of the transmission speed. Most epidemiological studies maintain constant transmission rates throughout the study period; such an approach proves both reasonable and efficient during the free transmission phase of the epidemic outbreak. However, the assumption of a fixed transmission rate becomes less valid once local governments implement response measures. It is anticipated that, following the initial free spread of the disease, the transmission rate will decrease over time as government control measures and public protection against the outbreak take effect. Considering that different transmission rates over time might solve this problem, it would also introduce unnecessary model complexity and potential overfitting concerns. A more parsimonious solution would be parameterizing the transmission rate as a continuous function of time.

Therefore, in analyzing large-scale epidemics, the model divides the outbreak into two distinct phases: the initial free transmission phase and the subsequent intervention phase, with particular emphasis on the latter. The free transmission phase is assumed to have a fixed transmission rate β_0, which also serves as the initial transmission rate for the intervention phase. As containment measures are implemented and progressively strengthened, the transmission rate in the intervention phase gradually decreases over time. This dynamic reduction in transmission probability directly modulates the effective reproductive number R, where R represents the average number of susceptible individuals who acquire infection from a single infectious case within a susceptible population. Through this mechanism, the model achieves the public health objective of epidemic containment via transmission suppression.

On the basis of the above description, the VCSEIR model $\mathcal{M}$ is set as follows:

$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= -\frac{\beta_0 S I^r \tau^t}{N} - \frac{\mu\beta_0 S I^\mu \tau^t}{N},\\
\frac{\mathrm{d}E}{\mathrm{d}t} &= \frac{\beta_0 S I^r \tau^t}{N} + \frac{\mu\beta_0 S I^\mu \tau^t}{N} - \frac{E}{Z},\\
\frac{\mathrm{d}I^r}{\mathrm{d}t} &= \alpha\frac{E}{Z} - \frac{I^r}{D},\\
\frac{\mathrm{d}I^\mu}{\mathrm{d}t} &= (1-\alpha)\frac{E}{Z} - \frac{I^\mu}{D}.
\end{aligned} \tag{1}$$

The model preserves S and E as the number of people in the susceptible and exposed groups, which is consistent with the classic model. I^r and I^μ are the confirmed population and unconfirmed population respectively. N denotes the total regional population. β_0 represents the initial transmission rate, which is set as a constant. The model sets parameter τ as the attenuation rate quantifying the daily reduction in transmission probability; such a setting gives the transmission rate of a confirmed case at time t as $\beta_0\tau^t$. α represents the case confirmation rate among infected individuals, and μ represents the relative reduction rate in the transmission rate of unconfirmed cases relative to that of confirmed cases. Z and D are the mean latent period and the average infectious duration of cases, respectively.

The solutions of the above simultaneous nonlinear equations are obtained via the fourth-order Runge–Kutta method. Notably, the right-hand side terms in Eq. (1) are numerically implemented through Poisson-distributed sampling, following established methodologies in epidemic modeling[3, 22]. The Poisson sampling mechanism was designed not only to capture the discrete nature of disease transmission events but also to provide robustness to the numerical solutions, making the model better able to cope with potential data fluctuations and measurement uncertainties.

Next, we briefly introduce the Ensemble Adjustment Kalman Filter (EAKF) algorithm[23, 24], which was used to estimate the unobservable variables of the model.

As defined in Ref. [21], the unobservable variables in the model can be categorized into two components: the state variables $\boldsymbol{X}=(S,E,I^r,I^\mu)^\top$, and the global parameters $\theta=(\alpha,\tau,\mu,Z,D)^\top$. These components exhibit different temporal characteristics: while the state variables dynamically evolve throughout the epidemic timeline, the global parameters remain time-invariant. Therefore, although independent estimates are obtained for both the state variables and the global parameters at each time point t, the final estimation of the global parameters is achieved through temporal averaging across all time points.

The core idea of EAKF is to assume that the prior and likelihood of the parameters are in accordance with the Gaussian distribution. After generating a certain number of ensemble members as samples, the Bayesian rule is applied to these ensemble members to obtain the posterior estimation, and then the algorithm sets the mean of posterior estimation over the ensemble members as the prior estimation of the next stage and iterates until the result converges. Since all state vari-

CSTR: 32290.14.JUSTC-2025-0014
DOI: 10.52396/JUSTC-2025-0014
JUSTC, **2025**, 55(X):

ables and global parameters in our framework are inherently unobservable, in the actual operation, we need to use the observations, i.e. daily confirmed case counts, to participate in the iteration. Specifically, for the observation o, we have

$$o_{t,\mathrm{post}}^{(\ell)} = \frac{\sigma_{t,\mathrm{obs}}^2}{\sigma_{t,\mathrm{obs}}^2+\sigma_{t,\mathrm{prior}}^2}\bar{o}_{t,\mathrm{prior}} + \frac{\sigma_{t,\mathrm{prior}}^2}{\sigma_{t,\mathrm{obs}}^2+\sigma_{t,\mathrm{prior}}^2}o_{t,\mathrm{obs}} + \sqrt{\frac{\sigma_{t,\mathrm{obs}}^2}{\sigma_{t,\mathrm{obs}}^2+\sigma_{t,\mathrm{prior}}^2}}\left(o_{t,\mathrm{prior}}^{(\ell)}-\bar{o}_{t,\mathrm{prior}}\right). \tag{2}$$

Here, $o_{t,\mathrm{prior}}^{(\ell)}$ is the prior estimation of the ℓth ensemble member at time t, $\bar{o}_{t,\mathrm{prior}} = \frac{\sum_{\ell=1}^{L} o_{t,\mathrm{prior}}^{(\ell)}}{L}$, and $o_{t,\mathrm{post}}^{(\ell)}$ is the corresponding posterior estimation. $\sigma_{t,\mathrm{obs}}^2$ and $\sigma_{t,\mathrm{prior}}^2$ are the variances set for the observations and the prior estimation, respectively. Next, for any unobservable variable w, let

$$w_{t,\mathrm{post}}^{(\ell)} = w_{t,\mathrm{prior}}^{(\ell)} + \frac{\varsigma}{\sigma_{t,\mathrm{prior}}^2}\left(o_{t,\mathrm{post}}^{(\ell)} - o_{t,\mathrm{prior}}^{(\ell)}\right), \tag{3}$$

where $w_{t,\mathrm{prior}}^{(\ell)}$ and $w_{t,\mathrm{post}}^{(\ell)}$ are the prior and posterior estimation for the ℓth ensemble member at time t, and ς is the sample covariance of $\{w_{t,\mathrm{prior}}\}$ and $\{o_{t,\mathrm{prior}}\}$. More details of EAKF can be found in Refs. [3, 21, 23], so we omit them.

Finally, we explain the reporting delay introduced into the model. Empirical epidemic data reveal that there are inherent delays between case confirmation and official reporting. Therefore, we formally define the "reporting delay" t_d as the temporal interval spanning from case confirmation to inclusion. In the inference of real data, we shift the new cases reported each day back by t_d days before including them in the daily reported case count. Following the established methodology in Ref. [23], we parameterize the reporting delay distribution via a gamma distribution, where its parameters are empirically predefined.

3 Theoretical results

3.1 Identifiability

In this section, we establish the theoretical foundation of our model. The first property we consider is the identifiability of the model structure. Before discussing our model, we first introduce the definition of identifiability for a dynamical system.

Let $\mathbf{X}(t) = \{X_1(t), \cdots, X_K(t)\}^{\mathrm{T}}$ be a set of n-dimensional variables, where K is the number of variables in $\mathbf{X}(t)$ (noticed that $K=4$ in our model). Then, consider the ordinary differential equation model with unknown parameters

$$\begin{cases} \mathbf{X}(t)' = \mathbf{F}[t, \mathbf{X}(t), \boldsymbol{\theta}], & \forall t \in [t_1, t_n]; \\ \mathbf{X}(t_1) = \mathbf{x}_1, \end{cases} \tag{4}$$

where $\boldsymbol{\theta} = \{\theta_1(t), \cdots, \theta_d(t)\}^{\top}$ represents a set of unknown parameters, $\mathbf{F}$ is a vector of known differentiable functions, and $\mathbf{X}(t_1) = \mathbf{x}_1$ is the initial value of the variables. We claim that the model structure is globally identifiable at $\boldsymbol{\theta}^*$ if the equations in $\boldsymbol{\theta}$ that arise from the equivalence

$$\mathbf{X}(t,\boldsymbol{\theta}) \equiv \mathbf{X}(t,\boldsymbol{\theta}^*) \quad \forall t \tag{5}$$

have the only solution $\boldsymbol{\theta} = \boldsymbol{\theta}^*$.

The global identifiability is defined as above, but it is often difficult to satisfy in practice. Therefore, we introduce the local identifiability as follows. For Eq. (5), if $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ results when $\boldsymbol{\theta}$ is restricted to an open neighborhood of $\boldsymbol{\theta}^*$, then the model structure is locally identifiable.

Now, for the VCSEIR model, we present the conclusion of local identifiability as follows.

Theorem 1. If the derivatives of the state variables (S, E, I^r, I^μ) exist and are continuous, then the VCSEIR model is locally identifiable.

3.2 Asymptotic property

Next, we consider the asymptotic properties of the model, which include consistency and asymptotic normality. The establishment of consistency ensures that the difference between the parameter estimates and the real parameters decreases as the sample size increases. In other words, as the sample size increases, the parameter estimates converge to the true values. On the other hand, the establishment of asymptotic normality guarantees that the parameter estimates converge to the real parameters in a normal distribution form, thereby improving the accuracy and reliability of the parameter estimation.

Define $X_1(t) = S(t)$, $X_2(t) = E(t)$, $X_3(t) = I^r(t)$, $X_4(t) = I^\mu(t)$, then $\mathbf{X}(t) = (X_1(t), X_2(t), X_3(t), X_4(t))$. Note that in the VCSEIR model, the real number of people in each group $\mathbf{Y}(t) = (Y_1(t), Y_2(t), Y_3(t), Y_4(t))$ is not directly determined by the state variables $\mathbf{X}(t)$, but rather is determined via a random sample from a Poisson distribution, i.e.,

$$f(Y_k(t_i) \mid X_k(t_i)) = \frac{X_k(t_i)^{Y_k(t_i)} \mathrm{e}^{-X_k(t_i)}}{Y_k(t_i)!},$$
$$i = 1, 2, \cdots, n, \quad k = 1, 2, \cdots, K,$$

where the number of state variables $K = 4$. The measurement error $\mathcal{E}(t) = \mathbf{Y}(t) - \mathbf{X}(t) = (\mathcal{E}_1(t), \mathcal{E}_2(t), \mathcal{E}_3(t), \mathcal{E}_4(t))$ are independent and have a mean of 0.

In practical modeling, when the number of observations is large, the data provides more information than the prior distribution does, and the parameter estimation obtained by the EAKF algorithm is approximately the result obtained via maximum likelihood estimation. Therefore, we consider only the properties of the maximum likelihood estimation of the parameters. The theoretical foundation for the estimation of the maximum likelihood for similar models has been relatively well developed, and our work mainly refers to the proof in Refs. [22, 25].

Denote the parameter vector as $\boldsymbol{\theta} = (\alpha, \tau, \mu, Z, D)^{\top}$ and satisfy $\boldsymbol{\theta} \in \Theta_b$, where $b = 5$ is the number of global parameters and Θ_b is the parameter space. Maximizing the likelihood function is equivalent to minimizing the negative log-likelihood. Let $\tilde{\mathbf{X}}(t,\boldsymbol{\theta})$ be the numerical solution of $\mathbf{X}(t,\boldsymbol{\theta})$ given by $\boldsymbol{\theta}$ from the numerical method, then:

$$\begin{aligned} \hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\Theta_b} \left[-\ln \prod_{i=1}^{n} \prod_{k=1}^{K} f\left(Y_k(t_i) \mid \tilde{X}_k(t_i,\boldsymbol{\theta})\right) \right] = \\ \arg\min_{\boldsymbol{\theta}\in\Theta_b} \left[-\sum_{i=1}^{n} \sum_{k=1}^{K} \ln \frac{\mathrm{e}^{-\tilde{X}_k(t_i,\boldsymbol{\theta})} \tilde{X}_k(t_i,\boldsymbol{\theta})^{Y_k(t_i)}}{Y_k(t_i)!} \right] = \\ \arg\min_{\boldsymbol{\theta}\in\Theta_b} \sum_{i=1}^{n} \sum_{k=1}^{K} \left[\tilde{X}_k(t_i,\boldsymbol{\theta}) - Y_k(t_i) \ln \tilde{X}_k(t_i,\boldsymbol{\theta}) + \ln Y_k(t_i)! \right]. \end{aligned} \tag{6}$$

CSTR: 32290.14.JUSTC-2025-0014
DOI: 10.52396/JUSTC-2025-0014
JUSTC, **2025**, 55(X):

Assuming that $\mathbf{X}$ is continuously differentiable with respect to $\boldsymbol{\theta}$ and is not 0, solving Eq. (6) can be equivalent to solving Eq. (7).

$$\frac{\partial \sum_{i=1}^{n}\sum_{k=1}^{K}\left[\tilde{X}_k(t_i,\boldsymbol{\theta})-Y_k(t_i)\ln\tilde{X}_k(t_i,\boldsymbol{\theta})+\ln Y_k(t_i)!\right]}{\partial\boldsymbol{\theta}}=0, \tag{7}$$

which is equivalent to

$$\sum_{i=1}^{n}\sum_{k=1}^{K}\left[1-\frac{Y_k(t_i)}{\tilde{X}_k(t_i,\boldsymbol{\theta})}\right]\frac{\partial\tilde{X}_k(t_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=0. \tag{8}$$

Then, we can transform the problem into solving

$$\sum_{i=1}^{n}\sum_{k=1}^{K}\left[\tilde{X}_k(t_i,\boldsymbol{\theta})-Y_k(t_i)\right]\frac{\partial\tilde{X}_k(t_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=0. \tag{9}$$

As a consequence, we can approximately consider that

$$\hat{\boldsymbol{\theta}}=\arg\min_{\theta\in\Theta_b}\sum_{i=1}^{n}\sum_{k=1}^{K}\left[\tilde{X}_k(t_i,\boldsymbol{\theta})-Y_k(t_i)\right]^2. \tag{10}$$

The problem is subsequently transformed into a nonlinear least squares theoretical property analysis with time-varying coefficients. Before discussing the problem in detail, we first present some assumptions. Define $I=[t_1,t_n]$ as the time period we consider. Let E represent the expectation on both t and k, and let E_t be the expectation on t. We give the following.

Assumption 1. $\boldsymbol{\theta}\in\mathcal{A}$, where $\mathcal{A}$ is a compact subset in R^b with a diameter of R_θ.

Assumption 2. $\Omega=\{\mathbf{X}(t,\boldsymbol{\theta}):t\in I,\boldsymbol{\theta}\in\mathcal{A}\}$ is a closed bounded convex subset in R^K.

Assumption 3. For any $t\in I$, $c_1\leqslant\mathbf{Y}(t)\leqslant c_2$, where c_1 and c_2 are two constants satisfying $-\infty<c_1<c_2<+\infty$.

Assumption 4. The numerical method for solving Eq. (1) is of order p.

Assumption 5. All the p-order partial derivatives of $\mathrm{F}(t,\mathbf{X},\boldsymbol{\theta})$ with respect to t and $\mathbf{X}$ exist and are continuous.

Assumption 6. For any $\boldsymbol{\theta}\in\mathcal{A}$, $\mathrm{E}_t[\mathbf{X}(t,\boldsymbol{\theta})-\mathbf{X}(t,\boldsymbol{\theta}_0)]^2=0$ if and only if $\boldsymbol{\theta}=\boldsymbol{\theta}_0$.

Assumption 7. $\dfrac{\partial\mathbf{X}(t,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$ and $\dfrac{\partial^2\mathbf{X}(t,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}$ exist and are continuous, and are uniformly bounded for all $t\in I$ and $\boldsymbol{\theta}\in\mathcal{A}$.

Assumption 8. For the numerical solution $\tilde{\mathbf{X}}(t,\boldsymbol{\theta})$, $\dfrac{\partial\tilde{\mathbf{X}}(t,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$, and $\dfrac{\partial^2\tilde{\mathbf{X}}(t,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}$ exist and are continuous, and are uniformly bounded for all $t\in I$ and $\boldsymbol{\theta}\in\mathcal{A}$.

Assumption 9. Let $0<c_3<c_4<\infty$ be two constants. Let the joint density function of $(t,\mathbf{Y})$ be $\phi(t,\mathbf{y})$, then for any $(t,\mathbf{y})\in[t_1,t_n]\times[\mathbf{c}_1,\mathbf{c}_2]$, $c_3\leqslant\phi(t,\mathbf{y})\leqslant c_4$.

Assumption 10. The real parameter $\boldsymbol{\theta}_0$ is an interior point of $\mathcal{A}$.

Assumption 11. $V_1=\left\{E_t\left(\dfrac{\partial\mathrm{X}}{\partial\boldsymbol{\theta}}\dfrac{\partial\mathrm{X}}{\partial\boldsymbol{\theta}^{\mathrm{T}}}\right)\right\}^{-1}E_t\left(\dfrac{\partial\mathrm{X}}{\partial\boldsymbol{\theta}}\sum\dfrac{\partial\mathrm{X}}{\partial\boldsymbol{\theta}^{\mathrm{T}}}\right)\cdot\left\{E_t\left(\dfrac{\partial\mathrm{X}}{\partial\boldsymbol{\theta}}\dfrac{\partial\mathrm{X}}{\partial\boldsymbol{\theta}^{\mathrm{T}}}\right)\right\}^{-1}$ is positive definite at $\boldsymbol{\theta}_0$, where $\mathrm{E}_t[g(t)]$ is the expectation of the function $g(t)$.

Assumptions 1–5 are general requirements for the existence of numerical solutions to ordinary differential equation models. Since we use the fourth-order Runge–Kutta method to obtain the solution, the order of the numerical method is actually 4. Assumption 6 is the condition for identifiability and consistency. Since we have already proven that the parameter is locally identifiable at $\boldsymbol{\theta}_0$ in Theorem 1, the condition is sufficient for Assumption 6. Assumptions 7–10 are the conditions required for consistency and Assumption 11 is an extra condition required for asymptotic normality. These assumptions are common assumptions about the theoretical basis of ODE models, and similar assumptions can be found in Ref. [25].

On the basis of the above assumption, the consistency of the model parameter estimation can be established. We give:

Theorem 2. If there exists a constant $\lambda>0$ such that the step length of the solution is $s=O(n^{-\lambda})$, then under Assumptions 1–10, $\hat{\boldsymbol{\theta}}$ converges in probability to $\boldsymbol{\theta}_0$.

Next, we further consider the asymptotic normality of the model parameter estimation:

Theorem 3. If a constant $\lambda>1$ exists such that the step length of the solution is $s=O(n^{-\lambda})$, assuming that Assumptions 1–11 are satisfied, we have

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\xrightarrow{\mathrm{d}}N(0,\mathbf{V}),$$

where $\mathbf{V}=\mathrm{E}[\mathbf{Y}(t)-\mathbf{X}(t,\boldsymbol{\theta}_0)]^2\left\{\mathrm{E}_t\left[\dfrac{\partial X(t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_0}\right]\mathrm{E}_t\left[\dfrac{\partial X(t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_0}\right]^{\mathrm{T}}\right\}^{-1}$.

At this point, we have presented the consistency and asymptotic normality of the model estimates.

4 Real data application

In this section, to demonstrate the performance of the VCSEIR model, we used the model to analyze the 2020 COVID-19 data from different countries and regions. The daily COVID-19 case data for all analyzed regions were sourced from the World Health Organization (WHO) COVID-19 Dashboard (https://covid19.who.int).

Different countries in Asia experienced several outbreaks in 2020, during which local governments implemented strict control measures. For the COVID-19 data from Asia, we selected three cities—Beijing, China; Seoul, South Korea; and Tokyo, Japan—to illustrate the inference results and show the model performance. In contrast, European nations predominantly released nationally aggregated datasets, necessitating country-scale analyses. Accordingly, we utilized nationwide data from Germany and Italy, two European countries that experienced significant pandemic waves during the study period.

For all the areas considered, we set the initial transmission rate to be 1.12, which is the same as the transmission rate estimated in the free transmission phase (consistent with Ref. [3]). Since the VCSEIR model focuses primarily on the period following the implementation of government control measures, we list the timelines of these measures in the selected regions. The Chinese, Japanese, and South Korean governments imposed strict control measures on January 23, April 7, and August 16, 2020, respectively, in response to outbreaks in these countries. For the European outbreak in the first quarter of 2020, Germany and Italy implemented strict control measures on March 16 and March 9, respectively.

CSTR: 32290.14.JUSTC-2025-0014
DOI: 10.52396/JUSTC-2025-0014
JUSTC, **2025**, 55(X):

During the inference process, the base value of the transmission rate for the attenuation stage is set to 1.12, which is consistent with the transmission rate of the free spread stage reported in previous studies[3]. Table 1 presents the global parameter estimates for all the regions and countries under consideration.

Table 1 shows that the estimated attenuation rates for Germany and Italy are significantly higher than those for the regions in Asia. This suggests that the control measures implemented by local governments in Europe required more time to take effect, resulting in prolonged outbreaks. In fact, outbreaks in 2020 lasted longer in European countries than across Asia.

In addition, under the framework of the VCSEIR model, if Assumptions 1–11 hold, we can approximate the variance of the parameter estimates via Theorem 3. Importantly, in real data applications, both the real parameter $\boldsymbol{\theta}_0$ and the latent state variables $\mathbf{Y}(t)$ are unobservable. Therefore, we use the estimated parameter $\hat{\boldsymbol{\theta}}$ to represent $\boldsymbol{\theta}_0$. For $k=1,2,3,4$. The state variable gap $Y_k(t)-X_k(t)$ is estimated on the basis of the difference between the observed and estimated daily new confirmed cases, $o_t-\hat{o}_t$. The final estimated variance is also shown in Table 1, which can serve as a reference for the effectiveness of parameter estimation. Notably, since the epidemic data for Germany and Italy span a longer time period, the larger sample sizes result in smaller variances of the parameter estimates than those of other regions.

Using the inferred parameters in Table 1, we can estimate the daily case numbers within the framework of the VCSEIR model. In Fig. 1, we present the fitted curve of the daily reported count estimated compared with the actual observed values. Fig. 2 displays the corresponding fitting curve for the

Table 1. Estimated parameters for selected countries and regions via the VCSEIR model. The variances of the parameter estimates obtained according to Theorem 3 are shown in parentheses.

City or nation	$\hat{\tau}$	$\hat{\mu}$	$\hat{Z}$	$\hat{\alpha}$	$\hat{D}$
Beijing	0.719(0.0005)	0.505(0.0704)	3.346(0.0585)	0.472(0.0039)	3.265(0.1595)
Seoul	0.709(0.0007)	0.536(0.0655)	3.158(0.0713)	0.569(0.0054)	2.894(0.1528)
Tokyo	0.689(0.0007)	0.532(0.0673)	3.089(0.0775)	0.552(0.0054)	3.061(0.2243)
Germany	0.794(0.0002)	0.532(0.0574)	3.239(0.0333)	0.463(0.0024)	3.071(0.0883)
Italy	0.783(0.0002)	0.576(0.0566)	3.125(0.0171)	0.569(0.0013)	3.144(0.0523)

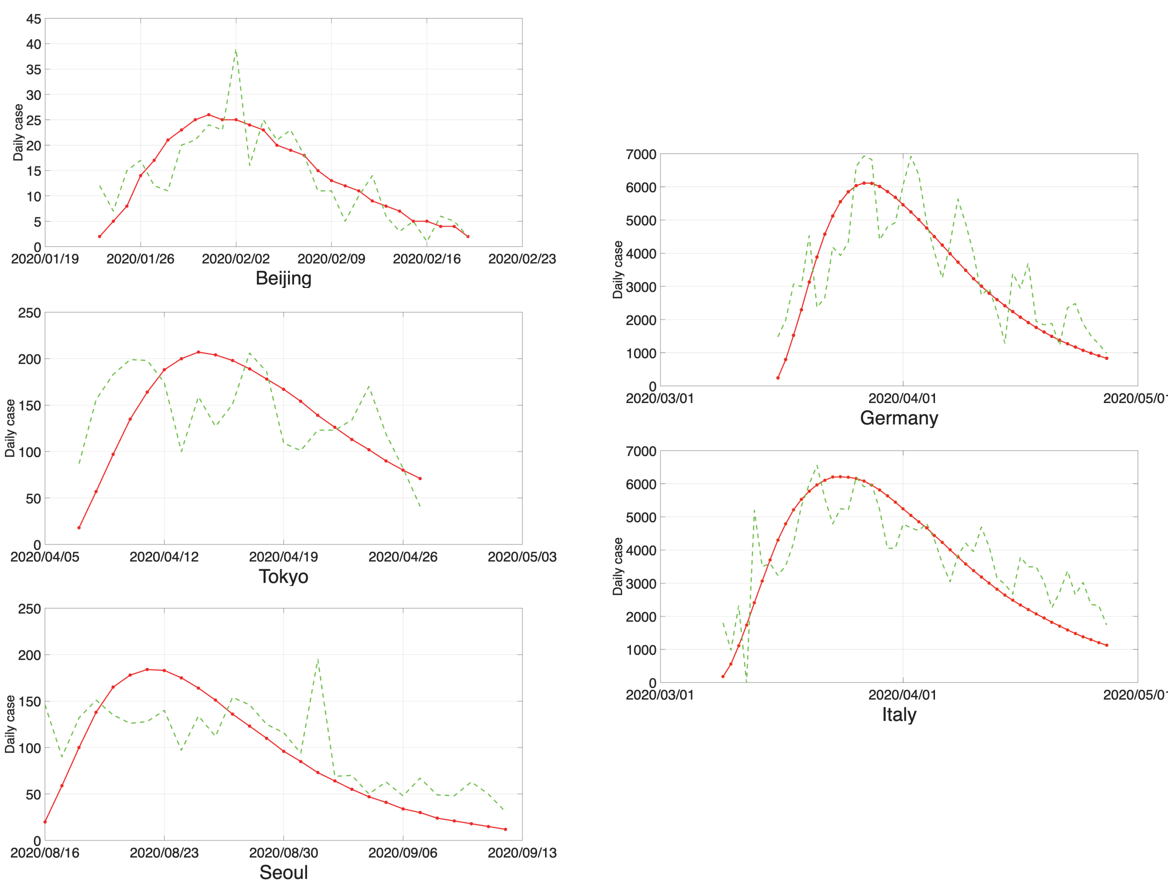


Fig. 1. The daily reported case count for different countries and regions (dotted green line) and the fitted curves of the daily reported count estimated by the VCSEIR model (solid red line).

CSTR: 32290.14.JUSTC-2025-0014
DOI: 10.52396/JUSTC-2025-0014
JUSTC, **2025**, 55(X):

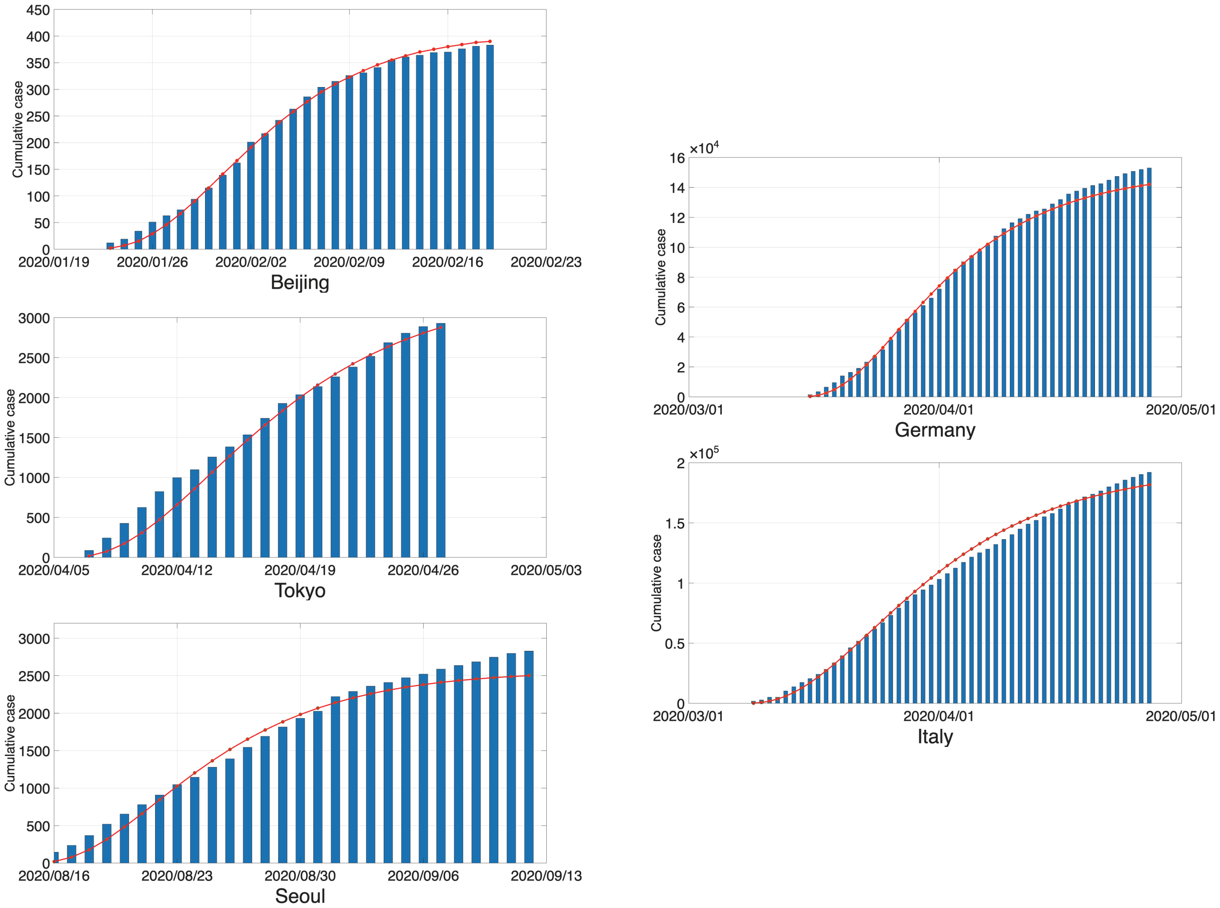


Fig. 2. The cumulative reported case counts over time for different countries and regions (bar) and the corresponding fitted curves of the cumulative reported count estimated by the VCSEIR model (solid red line).

cumulative number of confirmed cases. Figs. 1 and 2 show that the daily case numbers estimated by the model closely align with the real data. These results highlight the strong performance and reliability of the VCSEIR model.

5 Conclusions

In this paper, we refine the VCSEIR model within the framework of epidemic inference by establishing a rigorous theoretical foundation. This includes the identification of the model parameters, as well as the consistency and asymptotic normality of parameter estimation. Additionally, we have demonstrated the fitting performance of the VCSEIR model on data from different regions. These achievements ensure that the parameters of the model are not misidentified and increase the accuracy of parameter estimation, which is crucial for further statistical inference of the model parameters.

In summary, our work contributes to the extrapolation and estimation of the development of the attenuation stage of the epidemic. However, it should be noted that our model assumes that the parameters in the attenuation stage remain constant over time. While this assumption simplifies the model, it may lead to some deviation from the actual situation, as local governments may implement more stringent control measures over time. Detecting the change points of parameters in the attenuation stage could help address this issue. Similar approaches can be found in Ref. [26]. Precisely detecting change points in pandemic control measures will be a key focus of future research. On the other hand, aside from confirmed infection status, factors such as the age, gender, and other characteristics of infected patients may influence their transmission rate. Further subgroup stratification of the affected population could contribute to more accurate pandemic modeling, and this will be another important direction for further investigation.

Supporting information

The supporting information for this article can be found online at https://doi.org/10.52396/JUSTC-2025-0014. It includes corresponding proofs of all the theorems in the article.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (72111530199, 72293573).

Conflict of interest

The authors declare that they have no conflict of interest.

Biographies

Tianyi Sun is currently a Ph.D. student under the tutelage of Prof. Baisuo Jin at the University of Science and Technology of China. His research mainly focuses on the statistical analysis of dynamic network data and epidemiological models.

Baisuo Jin is a Professor at the University of Science and Technology of China (USTC). He received his Ph.D. degree from USTC in 2006. His research mainly focuses on large-dimensional stochastic matrices, experimental design, model selection, change points, and spatial statistics.

References

[1] Kermack W O, McKendrick A G. Contributions to the mathematical theory of epidemics: V. Analysis of experimental epidemics of mouse-typhoid; a bacterial disease conferring incomplete immunity. *Epidemiology & Infection*, **1939**, *39* (3): 271–288.

[2] Hethcote H W. The mathematics of infectious diseases. *SIAM Review*, **2000**, *42* (4): 599–653.

[3] Li R, Pei S, Chen B, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, **2020**, *368* (6490): 489–493.

[4] Sun H, Qiu Y, Yan H, et al. Tracking reproductivity of COVID-19 epidemic in China with varying coefficient SIR model. *Journal of Data Science*, **2020**, *18* (3): 455–472.

[5] Wu J T, Leung K, Leung G M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet*, **2020**, *395* (10225): 689–697.

[6] Prem K, Liu Y, Russell T W, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *The Lancet Public Health*, **2020**, *5* (5): e261–e270.

[7] IHME COVID-19 Forecasting Team. Modeling COVID-19 scenarios for the united states. *Nature Medicine*, **2021**, *27* (1): 94–105.

[8] Gatto M, Bertuzzo E, Mari L, et al. Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. *Proceedings of the National Academy of Sciences*, **2020**, *117* (19): 10484–10491.

[9] Pak A, Adegboye O A, Adekunle A I, et al. Economic consequences of the COVID-19 outbreak: the need for epidemic preparedness. *Frontiers in Public Health*, **2020**, *8*: 241.

[10] Nalbandian A, Sehgal K, Gupta A, et al. Post-acute COVID-19 syndrome. *Nature Medicine*, **2021**, *27* (4): 601–615.

[11] Chakraborty I, Maity P. COVID-19 outbreak: Migration, effects on society, global environment and prevention. *Science of the Total Environment*, **2020**, *728*: 138882.

[12] Tian H, Liu Y, Li Y, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science*, **2020**, *368* (6491): 638–642.

[13] Jia J S, Lu X, Yuan Y, et al. Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature*, **2020**, *582* (7812): 389–394.

[14] Kissler S M, Tedijanto C, Goldstein E, et al. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*, **2020**, *368* (6493): 860–868.

[15] Gu J, Yan H, Huang Y, et al. Comparing containment measures among nations by epidemiological effects of COVID-19. *National Science Review*, **2020**, *7* (12): 1847–1851.

[16] Yan H, Zhu Y, Gu J, et al. Better strategies for containing COVID-19 pandemic: a study of 25 countries via a vSIADR model. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **2021**, *477* (2248): 20200440.

[17] Leung K, Wu J T, Liu D, et al. First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. *The Lancet*, **2020**, *395* (10233): 1382–1393.

[18] Tian T, Tan J, Luo W, et al. The effects of stringent and mild interventions for coronavirus pandemic. *Journal of the American Statistical Association*, **2021**, *116* (534): 481–491.

[19] Daszak P, Jones K. Global trends in emerging infectious diseases. *Nature*, **2008**, *451* (7181): 990–993.

[20] Morens D M, Folkers G K, Fauci A S. The challenge of emerging and re-emerging infectious diseases. *Nature*, **2004**, *430* (6996): 242–249.

[21] Sun T, Jin B, Wu Y, et al. A study of the attenuation stage of a global infectious disease. *Frontiers in Public Health*, **2024**, *12*: 1379481.

[22] Jiang Y. Research on interpretable modeling of dynamical systems. Thesis. Guangzhou, China: Sun Yat-sen University, **2024**.

[23] Pei S, Kandula S, Yang W, et al. Forecasting the spatial transmission of influenza in the united states. *Proceedings of the National Academy of Sciences*, **2018**, *115* (11): 2752–2757.

[24] Anderson J L. An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, **2001**, *129* (12): 2884–2903.

[25] Xue H, Miao H, Wu H. Sieve estimation of constant and time-varying coefficients in nonlinear ordinary differential equation models by considering both numerical error and measurement error. *Annals of Statistics*, **2010**, *38* (4): 2351–2387.

[26] Dehning J, Zierenberg J, Spitzner F P, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science*, **2020**, *369* (6500): eabb9789.

CSTR: 32290.14.JUSTC-2025-0014
DOI: 10.52396/JUSTC-2025-0014
JUSTC, **2025**, 55(X):