Statistics
in Medicine WILEY

Enhanced empirical likelihood estimation of incubation period of COVID-19 by integrating published information

Zhongfeng Jiang[1] | **Baoying Yang[2]** | **Jing Qin[3]** | **Yong Zhou[4,5]**

[1]Academy of Mathematics and System Sciences, Chinese Academy of Science, Beijing, China

[2]Department of Statistics, College of Mathematics, Southwest Jiaotong University, Chengdu, China

[3]National Institute of Allergy and Infectious Diseases, National Institute of Health, Bethesda, Maryland

[4]Key Laboratory of Advanced Theory and Application in Statistics and Data Science, MOE, Shanghai, China

[5]Academy of Statistics and Interdisciplinary Sciences, East China Normal University, Shanghai, China

Correspondence
Baoying Yang, Department of Statistics, College of Mathematics, Southwest Jiaotong University, Chengdu, Sichuan 611756, China.
Email: yangbaoying@swjtu.edu.cn

Since the outbreak of the new coronavirus disease (COVID-19), a large number of scientific studies and data analysis reports have been published in the *International Journal of Medicine and Statistics*. Taking the estimation of the incubation period as an example, we propose a low-cost method to integrate external research results and available internal data together. By using empirical likelihood method, we can effectively incorporate summarized information even if it may be derived from a misspecified model. Taking the possible uncertainty in summarized information into account, we augment a logarithm of the normal density in the log empirical likelihood. We show that the augmented log-empirical likelihood can produce enhanced estimates for the underlying parameters compared with the method without utilizing auxiliary information. Moreover, the Wilks' theorem is proved to be true. We illustrate our methodology by analyzing a COVID-19 incubation period data set retrieved from Zhejiang Province and summarized information from a similar study in Shenzhen, China.

KEYWORDS
augmented log-empirical likelihood, COVID-19, incubation period, meta-analysis, Wilks' theorem

1 | INTRODUCTION

Since the outbreak of coronavirus disease 2019 (COVID-19), it has been spreading rapidly around the world and becoming a global pandemic. We are facing an unprecedented challenge to contain this disease eruption. Coronavirus treatment and vaccine research are moving at a record speed. At the same time, a large number of scientific research and data analysis reports about COVID-19 have been published in major international medical and statistical journals. Pooling the existing COVID-19 research results together for protection against this disease has become an indispensable task for global scientists. In epidemiology, the distribution of the incubation period has important clinical significance, such as tracking the source of infection and the path of transmission and determining the period of medical observation or isolation. Unfortunately, due to lack of knowledge or judgment on exposure time, the incubation period is almost unobservable

or even observable but with interval censoring and selection bias. The symptomatic onset times for those confirmed individuals, however, can be easily ascertained. Based on the COVID-19 daily updates from provincial and municipal health commissions in China, Qin et al[1] have noticed that there is an abundance of cases who asymptomatically left Wuhan, the epicenter of COVID-19 in China, and developed symptoms outside Wuhan. They have defined a forward time as the elapse between departure from Wuhan and symptom onset time in somewhere else. They estimate the incubation period distribution indirectly by using the connection between forward time density and incubation period density (see (1) in Section 2). The forward time can also be treated as a truncated version of incubation period when the truncation variable has a uniform distribution, see, for example, Linton et al.[2]

In this article, the individual-level data are collected from municipal health committees in Zhejiang Province of China. Out of 1123 confirmed cases, there are 147 confirmed individuals who left Wuhan asymptomatically from 19 January 2020 to 23 January 2020 and later developed symptoms in Zhejiang Province. Forward times for those 147 confirmed cases are defined as the times between departure from Wuhan and symptom onset. The underlying parameters of forward time or incubation period of COVID-19 can be estimated straightforwardly following Qin et al.[1] However, there are many other available results about forward time or incubation period in medical and statistical literature. For example, MedRxiv is a public online site to post pre-published medical manuscripts. Bi et al[3] have published a timely paper to analyze confirmed cases identified between 14 January 2020 and 12 February 2020 in Shenzhen, China. They have used a Log-normal distribution to fit the observed forward times (the time from arrival to symptom onset) for 191 travelers developed symptoms after arriving in Shenzhen.

A natural question is how to incorporate summarized information with individual level data collected in Zhejiang province to get more efficient estimates for the distribution of forward time or incubation period. Meta-analysis[4-6] is a systematic way to utilize the summarized information from several relevant studies. In practice, comprehensive individual-level data may not be publicly available due to privacy concerns or other issues. Frequently, only limited individual-level data and summarized statistics are available. In this situation, a hybrid approach by combining meta-analysis results and individual-level data has been shown to produce more precise estimates of underlying parameters.[7]

Inspired by the idea of meta-analysis, we develop a low-cost method to integrate multiple external auxiliary summarized results and available individual level data by using Owen's empirical likelihood method.[8] When the external data sample size m is much larger than the internal data sample size n, the uncertainty of the auxiliary summarized information derived from the external study is negligible.[7] However, in our real data example, the magnitudes m and n are comparable, so the uncertainty of summarized information cannot be ignorable. To take the possible variability of summarized information into consideration, we treat it as an observation of a normal random variable. Furthermore, we augment the log-empirical likelihood by a logarithm of normal density derived from the auxiliary information. The augmented log-likelihood can produce enhanced estimates of the underlying parameters. We show that the Wilks' theorem also holds to be true.

This article is organized as follows. In Section 2, we present our methodology and large sample results. Simulation results are given in Section 3. In Section 4, we apply our proposed method to estimate the incubation period distribution of COVID-19 by integrating the Zhejiang Province individual forward time data and the external summarized information from Bi et al.[3] We conclude this article with some remarks in Section 5.

2 | METHODOLOGY

Let V be the duration from departure of Wuhan to the onset of symptoms. Denote A as the time elapse from disease onset to the departure of Wuhan. The incubation period is $T = A + V$. Due to uncertainty on the time of contracting the COVID-19, each patient's A is not available. Moreover, patients who had symptoms before their departure of Wuhan were not included in our data set. This is a truncation problem, see, for example, Qin et al.[7] If we assume A follows a uniform distribution, then the truncated version of A and V can be treated as backward time and forward time in the renewal process theory, respectively, see for example Cox.[9] Let $f(t, \boldsymbol{\eta})$ be the density of unobservable incubation period T, where the form of $f(\cdot)$ is known but $\boldsymbol{\eta}$ is an unknown vector parameter. By the renewal theory,[9] the forward time V has density

$$g(v) = \frac{\overline{F}(v)}{\mu}, \quad v \geq 0, \tag{1}$$

where $\overline{F}(x) = Pr(T \geq x) = \int_x^\infty f(t, \boldsymbol{\eta})dt$, and $\mu = \int_0^\infty t f(t, \boldsymbol{\eta})dt$ is the mean of T. Denote by $G(v)$ the corresponding cumulative distribution of $g(v, \boldsymbol{\eta})$ and

$$\boldsymbol{a}_1(v, \boldsymbol{\eta}) = \frac{\partial \log g(v, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}}. \tag{2}$$

Obviously, the score function in (2) satisfies

$$\boldsymbol{\psi}_1 = E(\boldsymbol{a}_1(v, \boldsymbol{\eta})) = \int \boldsymbol{a}_1(v, \boldsymbol{\eta}) \, dG(v) = \boldsymbol{0}.$$

In the presence of external studies, suppose summarized information from a similar study is available. It is derived from a density $h(v, \boldsymbol{\theta})$, which may or may not be the same as $g(v, \boldsymbol{\eta})$. Due to various reasons, the raw data from the external study are not available. The question is how to combine the existing data with the summarized information together for an enhanced inference on the underlying parameter $\boldsymbol{\eta}$. For example, Bi et al[3] have obtained summarized information under the assumption that the forward time follows a Log-normal distribution. Note that the Log-normal assumption may be misspecified since the forward time should have a monotone decreasing density.[7] Denote

$$\boldsymbol{a}_2(v, \boldsymbol{\theta}) = \frac{\partial \log h(v, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.$$

Following White,[10] under some regular conditions, no matter the density function $h(v, \boldsymbol{\theta})$ is correctly specified or not, we can conclude that there exists a $\boldsymbol{\theta}_*$ such that

$$\int \boldsymbol{a}_2(v, \boldsymbol{\theta}_*) \, dG(v) = 0$$

and the MLE $\hat{\boldsymbol{\theta}} \overset{P}{\to} \boldsymbol{\theta}_*$, where $\boldsymbol{\theta}_*$ is the unique minimum of Kullback-Leibler divergence

$$I(g, h, \boldsymbol{\theta}) = \int \log g(v)dG(v) - \int \log h(v, \boldsymbol{\theta})dG(v).$$

Motivated by this fact, we can construct an extra unbiased estimation equation

$$\boldsymbol{\psi}_2 = \int \boldsymbol{a}_2(v, \boldsymbol{\theta}_*) \, dG(v) = E(\boldsymbol{a}_2(v, \boldsymbol{\theta}_*)) = 0.$$

To combine above auxiliary information, we employ Owen's empirical likelihood method.[8] Denote the observed forward times as $\{v_1, \ldots, v_n\}$. If $\boldsymbol{\theta}_*$ is known, then the log empirical likelihood is

$$\sum_{i=1}^n \log w_i$$

subject to the constraints

$$\sum_{n=1}^n w_i = 1, \quad w_i \geq 0,$$

$$\sum_{i=1}^n w_i \boldsymbol{a}_1(v_i, \boldsymbol{\eta}) = 0, \quad \sum_{i=1}^n w_i \boldsymbol{a}_2(v_i, \boldsymbol{\theta}_*) = 0.$$

The corresponding profile empirical likelihood is $\ell(\boldsymbol{\eta}, \boldsymbol{\theta}_*, \lambda) = -\sum_{i=1}^n \log \{1 + \lambda^T \boldsymbol{a}(v_i, \boldsymbol{\eta}, \boldsymbol{\theta}_*)\}$, where $\boldsymbol{a}(v, \boldsymbol{\eta}, \boldsymbol{\theta}_*) = \left(\boldsymbol{a}_1^T(v, \boldsymbol{\eta}), \boldsymbol{a}_2^T(v, \boldsymbol{\theta}_*)\right)^T$, and the Lagrange multiplier λ is determined by

$$\sum_{i=1}^n \frac{\boldsymbol{a}(v_i, \boldsymbol{\eta}, \boldsymbol{\theta}_*)}{n\left(1 + \lambda^T \boldsymbol{a}(v_i, \boldsymbol{\eta}, \boldsymbol{\theta}_*)\right)} = 0.$$

However, $\boldsymbol{\theta}_*$ is unknown in general and has to be replaced by its estimator. If the external sample size m is comparable with the internal sample size n, then the uncertainty of $\widehat{\boldsymbol{\theta}}$ cannot be ignored.[7] The summarized information is usually reported as $\widehat{\boldsymbol{\theta}}$ with estimated covariance matrix $\widehat{\Sigma}$. Therefore $\widehat{\boldsymbol{\theta}}$ can be treated as a random variable generated from a normal distribution, that is,

$$\widehat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}_*, \widehat{\Sigma}).$$

This strategy has been used widely in meta-analysis literature, for example, DerSimonian and Laird.[11] As a consequence, we can augment the log empirical likelihood by

$$\ell_A(\boldsymbol{\eta}, \boldsymbol{\theta}, \lambda) = \ell(\boldsymbol{\eta}, \boldsymbol{\theta}, \lambda) - \frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})^T \widehat{\Sigma}^{-1}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})$$

$$= -\sum_{i=1}^{n} \log\left\{1 + \lambda^T \boldsymbol{a}(v_i, \boldsymbol{\eta}, \boldsymbol{\theta})\right\} - \frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})^T \widehat{\Sigma}^{-1}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}). \tag{3}$$

Then we maximize ℓ_A to get the solutions $\widetilde{\boldsymbol{\eta}}$ and $\widetilde{\boldsymbol{\theta}}$ as the estimation of $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$, respectively.

Under regular conditions C1–C3 stated in Appendix A, the consistency and the asymptotic normality of $\widetilde{\boldsymbol{\eta}}$ and $\widetilde{\boldsymbol{\theta}}$ are proven there. In detail, suppose that $\widehat{\Sigma}^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_*) \to N(\mathbf{0}, \boldsymbol{I})$ in distribution and $lim_{n \to \infty} m/n = c > 0$, then the proposed estimates $\widetilde{\boldsymbol{\eta}}$ and $\widetilde{\boldsymbol{\theta}}$ are consistent and $\sqrt{n}\left(\widetilde{\boldsymbol{\eta}}^T - \boldsymbol{\eta}_0^T, \widetilde{\boldsymbol{\theta}}^T - \boldsymbol{\theta}_*^T\right)^T$ asymptotically converges to a zero-mean normal distribution with covariance matrix

$$\Gamma = \begin{pmatrix} A_{13}A_{33}^{-1}A_{31} & A_{13}A_{33}^{-1}A_{32} \\ A_{23}A_{33}^{-1}A_{31} & A_{22} + A_{23}A_{33}^{-1}A_{32} \end{pmatrix}^{-1},$$

where $\boldsymbol{\eta}_0$ is the true value of parameter $\boldsymbol{\eta}$, and

$$A_{22} = \lim_{n \to \infty} \widehat{\Sigma}^{-1}/n, \quad A_{33} = E\left(\boldsymbol{a}(v_i, \boldsymbol{\eta}_0, \boldsymbol{\theta}_*)\boldsymbol{a}^T(v_i, \boldsymbol{\eta}_0, \boldsymbol{\theta}_*)\right),$$

$$A_{13} = A_{31}^T = E\left(\frac{\partial \boldsymbol{a}^T(v_i, \boldsymbol{\eta}_0, \boldsymbol{\theta}_*)}{\partial \boldsymbol{\eta}}\right), \quad A_{23} = A_{32}^T = E\left(\frac{\partial \boldsymbol{a}^T(v_i, \boldsymbol{\eta}_0, \boldsymbol{\theta}_*)}{\partial \boldsymbol{\theta}}\right).$$

Moreover, the terms $A_{22}, A_{33}, A_{13}, A_{23}$ can be consistently estimated, respectively, by

$$\widetilde{A}_{22} = \widehat{\Sigma}^{-1}\Big/n, \quad \widetilde{A}_{33} = \sum_{i=1}^{n} \frac{1}{n}\left(\boldsymbol{a}(v_i, \widetilde{\boldsymbol{\eta}}, \widetilde{\boldsymbol{\theta}})\boldsymbol{a}^T(v_i, \widetilde{\boldsymbol{\eta}}, \widetilde{\boldsymbol{\theta}})\right),$$

$$\widetilde{A}_{13} = \sum_{i=1}^{n} \frac{1}{n}\left(\frac{\partial \boldsymbol{a}^T(v_i, \widetilde{\boldsymbol{\eta}}, \widetilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\eta}}\right), \quad \widetilde{A}_{23} = \sum_{i=1}^{n} \frac{1}{n}\left(\frac{\partial \boldsymbol{a}^T(v_i, \widetilde{\boldsymbol{\eta}}, \widetilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right).$$

Furthermore, we can get the corresponding asymptotic covariance matrix estimation of $\sqrt{n}\left(\widetilde{\boldsymbol{\eta}}^T - \boldsymbol{\eta}_0^T, \widetilde{\boldsymbol{\theta}}^T - \boldsymbol{\theta}_*^T\right)^T$ or $\sqrt{n}\left(\widetilde{\boldsymbol{\eta}}^T - \boldsymbol{\eta}_0^T\right)$.

In contrast, the covariance matrix of empirical likelihood method without auxiliary information[7] is

$$\Gamma_2 = \begin{pmatrix} A_{13}A_{33}^{-1}A_{31} & A_{13}A_{33}^{-1}A_{32} \\ A_{23}A_{33}^{-1}A_{31} & A_{23}A_{33}^{-1}A_{32} \end{pmatrix}^{-1} = \left[\begin{pmatrix} A_{13} \\ A_{23} \end{pmatrix} A_{33}^{-1} \begin{pmatrix} A_{31} & A_{32} \end{pmatrix}\right]^{-1}.$$

If the external data sample size m is much smaller than the internal sample size n, then the improvement of our proposed method in parameter estimators may be negligible. However, when m is comparable with n, incorporating auxiliary information from external model does improve estimation efficiency. Specifically, if $\boldsymbol{a}(v, \boldsymbol{\eta}, \boldsymbol{\theta}) = \left(\boldsymbol{a}_1^T(v, \boldsymbol{\eta}), \boldsymbol{a}_2^T(v, \boldsymbol{\theta})\right)^T$, we have

$$\left[E\left(\frac{\partial \boldsymbol{a}_1^T(v_i, \boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}}\right)\right]^{-1} E\left(\boldsymbol{a}_1(v_i, \boldsymbol{\eta}_0)\boldsymbol{a}_1^T(v_i, \boldsymbol{\eta}_0)\right) \left[E\left(\frac{\partial \boldsymbol{a}_1(v_i, \boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}^T}\right)\right]^{-1} = \Gamma_2(1, 1) > \Gamma(1, 1).$$

where $\Gamma(1, 1)$, $\Gamma_2(1, 1)$ are the $(1, 1)$th submatrix of the 2×2 block matrix Γ and Γ_2, respectively. This illustrates that our estimator $\widetilde{\eta}$ is more efficient than the one without utilizing auxiliary information.

It is worth noting that the covariance matrix $\widehat{\Sigma}$ may be misspecified in practice. For example, White[10] showed the asymptotic normality of the quasi-maximum likelihood estimate even under the misspecified density function h:

$$\sqrt{m}(\widehat{\theta} - \theta_*) \xrightarrow{d} N(0, C(\theta_*)),$$

where matrix $C(\theta) = A(\theta)^{-1} B(\theta) A(\theta)^{-1}$, $A(\theta) = E\left[\partial^2 \log h(v_i, \theta)/\partial\theta \, \partial\theta^T\right]$ and $B(\theta) = E[\partial \log h(v_i, \theta)/\partial\theta \cdot \partial \log h(v_i, \theta)/\partial\theta^T]$, and $E(\cdot)$ is the expectation with respect to the true density function $g(v)$. Note that the covariance matrix $C(\theta_*)$ may be incorrectly estimated by $\widehat{A} = \frac{1}{m} \sum_{i=1}^{m} \partial^2 \log h(v_i, \widehat{\theta})/\partial\theta \, \partial\theta^T$ in practice. Similar to Liu et al,[5] we can show that our method is still effective and robust against misspecification of the covariance matrix. This robustness property greatly enhances the applicability of our proposed approach. In detail, denote Σ_W as a "working" covariance matrix of $\widehat{\theta}$, and $\widetilde{\eta}_W$, $\widetilde{\theta}_W$ as the new estimators obtained from (3) by replacing $\widehat{\Sigma}$ with Σ_W. If Σ_W^{-1}/n converges to a positive definite matrix A_W in probability as $n \to \infty$, then using the similar arguments in Appendix B, we can show that the new estimators $\widetilde{\eta}_W$, $\widetilde{\theta}_W$ are consistent, and $\sqrt{n}\left(\widetilde{\eta}_W^T - \eta_0^T, \widetilde{\theta}_W^T - \theta_*^T\right)^T$ asymptotically converges to a zero-mean normal distribution with covariance matrix

$$\Gamma_W = \begin{pmatrix} A_{13}A_{33}^{-1}A_{31} & A_{13}A_{33}^{-1}A_{32} \\ A_{23}A_{33}^{-1}A_{31} & A_W + A_{23}A_{33}^{-1}A_{32} \end{pmatrix}^{-1} \begin{pmatrix} A_{13}A_{33}^{-1}A_{31} & A_{13}A_{33}^{-1}A_{32} \\ A_{23}A_{33}^{-1}A_{31} & A_W(C/c)A_W + A_{23}A_{33}^{-1}A_{32} \end{pmatrix}$$

$$\times \begin{pmatrix} A_{13}A_{33}^{-1}A_{31} & A_{13}A_{33}^{-1}A_{32} \\ A_{23}A_{33}^{-1}A_{31} & A_W + A_{23}A_{33}^{-1}A_{32} \end{pmatrix}^{-1},$$

where C is the true asymptotic covariance matrix of $\sqrt{m}(\widehat{\theta} - \theta_*)$, and $c = lim_{n \to \infty} m/n$.

If interested in testing $\eta = \eta_1$, we can show that the Wilks' theorem holds to be true under assumptions C1–C3 specified in the Appendix. In detail, under the null hypothesis $H_0 : \eta = \eta_1$, the empirical likelihood ratio statistic

$$R_E(\eta_1) = 2\ell_A\left(\widetilde{\eta}, \widetilde{\theta}\right) - 2\ell_A\left(\eta_1, \widetilde{\theta}_1\right)$$

satisfies $R_E(\eta_1) \to \chi^2(q)$ in distribution, where $\widetilde{\theta}_1$ is the H_0 restricted MLE, that is, it maximizes $\ell_A(\eta_1, \theta)$ and q is the dimension of η.

3 | SIMULATION STUDY

In this section, we conduct extensive simulations to investigate the finite sample performance of the proposed method. For convenience of description, we define "PEL" as our proposed empirical likelihood method, and "CEL" as the classic empirical likelihood estimation without utilizing auxiliary summarized information from external data.

In the first simulation study, n independent and identically distributed observations are generated from the underlying distribution

$$g(v, \eta) = \frac{\overline{F}(v, \eta)}{\mu}, \quad v \geq 0,$$

where $\overline{F}(t)$ is the complementary to the cumulative distribution function of a Weibull distribution with density function $f(t) = \alpha\beta(t\beta)^{\alpha-1} \exp\{-(t\beta)^\alpha\}$, $\mu = \int_0^\infty tf(t)dt, t \geq 0$, and $\eta = (\alpha, \beta)$. Furthermore, we choose true $(\alpha, \beta) = (2, 2)$ or $(2, 0.5)$. In addition, m independent and identically distributed external samples are generated from the same distribution as the internal data. By mimicking Bi et al[3] and fitting a misspecified Log-normal distribution with density function

$$h(v) = \frac{1}{v\sqrt{2\pi b}} \exp\left\{-\frac{1}{2}\left(\frac{\ln(v) - a}{b}\right)^2\right\},$$

TABLE 1 Simulation results for integrated analysis for individual data and auxiliary summary information

Case	Methods	α				β			
		Mean	Bias	SD	Median	Mean	Bias	SD	Median
		$\alpha = 2$				$\beta = 2$			
$n = 200, m = 500$	PEL	2.0628	0.0628	0.2631	2.0215	1.9914	−0.0086	0.1101	1.9755
	CEL	2.0533	0.0533	0.2876	2.0351	2.0102	0.0102	0.1665	2.0031
$n = 500, m = 1000$	PEL	2.0526	0.0526	0.1471	2.0561	1.9855	−0.0145	0.0715	1.9764
	CEL	2.0515	0.0515	0.1646	2.0418	1.9901	−0.0099	0.1011	1.9828
		$\alpha = 2$				$\beta = 0.5$			
$n = 200, m = 500$	PEL	2.0743	0.0743	0.2476	2.0508	0.4974	−0.0026	0.0291	0.4937
	CEL	2.0856	0.0856	0.2736	2.0561	0.4988	−0.0012	0.0422	0.4976
$n = 500, m = 1000$	PEL	2.0144	0.0144	0.1435	1.9983	0.4998	−0.0002	0.0183	0.4988
	CEL	2.0123	0.0123	0.1594	2.0002	0.5015	0.0015	0.0264	0.5011

we obtain summarized external information such as the estimated unknown parameters $\theta = (a, b)$ and their corresponding covariance matrix.

We compare the performance of our proposed method PEL with CEL for $n = 200, m = 500$ or $n = 500, m = 1000$ over 200 replicates, and the simulation results are presented in Table 1. Table 1 reports the sample mean (Mean), sample bias (Bias), sample standard deviation (SD), and the sample median (Median) of the estimators. Overall, our proposed estimator performed well. The standard deviation of PEL estimator is smaller than that from the CEL estimator. This is expected since auxiliary information was used in the PEL method. In addition, the larger the sample size is, the more precise the PEL and CEL estimators are.

Motivated by the sensitivity analysis in Qin et al,[1] we conduct the second simulation to evaluate the performance of our method with a mixture density function. Note that among those patients considered in the real data analysis, a small portion of them might contract the disease on their way out of Wuhan. In such case, the observed forward times can be treated as a mixture of the "real" forward times and incubation periods. So in our second scenario, the observed individual data are generated from a mixture density function,

$$g(v, \boldsymbol{\eta}) = \pi f(v, \boldsymbol{\eta}) + (1 - \pi)\frac{\overline{F}(v, \boldsymbol{\eta})}{\mu},$$

where π is the proportion of newly infected COVID-19 patients from bus stations, train stations, airports, and so on. The incubation period is still specified as the Weibull distribution, that is, f, $\overline{F}$, and μ are the same as in the first simulation. We generate n and m samples as the observed individual level data and external data from the mixture model with varying but correctly specified $\pi = 0.05, \ 0.1, \ 0.2, \ 0.5$, respectively. The true value of (α, β) is set to be $(2, 2)$ or $(2, 0.5)$. We still use the Log-normal distribution to fit the external data and get summarized auxiliary information.

We compare our proposed method PEL with CEL for $n = 200, m = 500$ or $n = 500, m = 1000$ over 200 independent replicates. The Monte Carlo results are reported in Table 2. Similar to the simulation results in Table 1, the numerical performance of our method is pretty good, and the standard deviation of PEL estimator is smaller than that of the CEL estimator. This also verifies that our proposed method achieves efficiency gain.

Two key assumptions are made in the first two simulation studies, that is, π is given correctly and the incubation period density f is also correctly specified. To test the robustness of our method, we conducted additional sensitivity analysis by violating the two key assumptions. In the robustness analysis about the misspecification of π values, the observed data and external data were generated from the mixture density function

$$g(v, \boldsymbol{\eta}) = \pi_0 f(v, \boldsymbol{\eta}) + (1 - \pi_0)\frac{\overline{F}(v, \boldsymbol{\eta})}{\mu},$$

where $\pi_0 = 0.1$. The density function of incubation period f is chosen to be Weibull. When we analyze the observed data, the mixture proportion is chosen to be $\pi = 0, \ 0.05, \ 0.1, \ 0.2$, and 0.5, respectively. The summarized auxiliary information

TABLE 2 Simulation results under mixture density with varying correctly specified π

Case	Methods	α				β			
		Mean	Bias	SD	Median	Mean	Bias	SD	Median
		$\alpha = 2$				$\beta = 2$			
		$\pi = 0.05$							
$n = 200, m = 500$	PEL	2.0585	0.0585	0.2555	2.0080	1.9931	−0.0069	0.1011	1.9833
	CEL	2.0561	0.0561	0.2781	2.0205	2.0082	0.0082	0.1561	1.9956
$n = 500, m = 1000$	PEL	2.0263	0.0263	0.1482	2.0227	1.9975	−0.0025	0.0749	1.9970
	CEL	2.0299	0.0299	0.1614	2.0214	1.9989	−0.0011	0.0962	1.9951
		$\pi = 0.1$							
$n = 200, m = 500$	PEL	2.0657	0.0657	0.2464	2.0138	1.9908	−0.0092	0.0979	1.9843
	CEL	2.0633	0.0633	0.2779	2.0395	2.0002	0.0002	0.1464	1.9845
$n = 500, m = 1000$	PEL	2.0211	0.0211	0.1360	2.0183	1.9983	−0.0017	0.0679	1.9985
	CEL	2.0244	0.0244	0.1524	2.0209	2.0005	0.0005	0.0947	1.9957
		$\pi = 0.2$							
$n = 200, m = 500$	PEL	2.0455	0.0455	0.2085	2.0191	1.9932	−0.0068	0.0856	1.9906
	CEL	2.0501	0.0501	0.2368	2.0242	1.9986	−0.0014	0.1242	1.9905
$n = 500, m = 1000$	PEL	2.0242	0.0242	0.1134	2.0195	1.9935	−0.0065	0.0542	1.9898
	CEL	2.0231	0.0231	0.1238	2.0213	2.0008	0.0008	0.0731	1.9981
		$\pi = 0.5$							
$n = 200, m = 500$	PEL	2.0272	0.0272	0.1396	2.0067	1.9939	−0.0061	0.0621	1.9920
	CEL	2.0360	0.0360	0.1597	2.0297	1.9922	−0.0078	0.1063	1.9895
$n = 500, m = 1000$	PEL	2.0135	0.0135	0.0901	2.0152	1.9955	−0.0045	0.0399	1.9971
	CEL	2.0111	0.0111	0.0961	2.0065	2.0030	0.0030	0.0638	2.0061
		$\alpha = 2$				$\beta = 0.5$			
		$\pi = 0.05$							
$n = 200, m = 500$	PEL	2.0676	0.0676	0.2428	2.0651	0.4981	−0.0019	0.0279	0.4943
	CEL	2.0763	0.0763	0.2647	2.0622	0.4998	−0.0002	0.0401	0.4984
$n = 500, m = 1000$	PEL	2.0243	0.0243	0.1308	2.0122	0.4988	−0.0012	0.0168	0.4979
	CEL	2.0293	0.0293	0.1453	2.0261	0.4987	−0.0013	0.0237	0.4986
		$\pi = 0.1$							
$n = 200, m = 500$	PEL	2.0761	0.0761	0.2314	2.0568	0.4976	−0.0024	0.0261	0.4958
	CEL	2.0851	0.0851	0.2551	2.0922	0.4983	−0.0017	0.0382	0.4954
$n = 500, m = 1000$	PEL	2.0229	0.0229	0.1211	2.0245	0.4991	−0.0009	0.0152	0.4991
	CEL	2.0251	0.0251	0.1338	2.0232	0.4993	−0.0007	0.0223	0.4976
		$\pi = 0.2$							
$n = 200, m = 500$	PEL	2.0608	0.0608	0.2064	2.0565	0.4991	−0.0009	0.0223	0.4981
	CEL	2.0629	0.0629	0.2231	2.0384	0.5004	0.0004	0.0335	0.4972
$n = 500, m = 1000$	PEL	2.0318	0.0318	0.1275	2.0275	0.4976	−0.0024	0.0148	0.4976
	CEL	2.0344	0.0344	0.1351	2.0363	0.4976	−0.0024	0.0201	0.4972
		$\pi = 0.5$							
$n = 200, m = 500$	PEL	2.0255	0.0255	0.1423	2.0212	0.5003	0.0003	0.0157	0.5007
	CEL	2.0355	0.0355	0.1591	2.0254	0.4995	−0.0005	0.0247	0.4966
$n = 500, m = 1000$	PEL	2.0179	0.0179	0.0855	2.0184	0.4985	−0.0015	0.0101	0.4995
	CEL	2.0203	0.0203	0.0954	2.0228	0.4985	−0.0015	0.0158	0.4993

TABLE 3 Simulation results for sensitive analysis with the misspecification of π value, and the true value $\pi_0 = 0.1$

Case	Methods	$\alpha = 2$				$\beta = 0.5$			
		Mean	Bias	SD	Median	Mean	Bias	SD	Median
		$\pi = 0$							
$n = 200, m = 500$	PEL	2.1854	0.1854	0.2690	2.1488	0.4616	−0.0384	0.0246	0.4616
	CEL	2.1666	0.1666	0.2840	2.1630	0.4678	−0.0322	0.0365	0.4628
$n = 500, m = 1000$	PEL	2.1471	0.1471	0.1585	2.1320	0.4630	−0.0370	0.0156	0.4638
	CEL	2.1425	0.1425	0.1777	2.1279	0.4647	−0.0353	0.0229	0.4635
		$\pi = 0.05$							
$n = 200, m = 500$	PEL	2.1308	0.1308	0.2519	2.0919	0.4790	−0.0210	0.0242	0.4799
	CEL	2.1249	0.1249	0.2650	2.1088	0.4825	−0.0175	0.0364	0.4796
$n = 500, m = 1000$	PEL	2.0845	0.0845	0.1492	2.0671	0.4817	−0.0183	0.0157	0.4819
	CEL	2.0898	0.0898	0.1704	2.0721	0.4820	−0.0180	0.0235	0.4810
		$\pi = 0.1$							
$n = 200, m = 500$	PEL	2.0606	0.0606	0.2342	2.0341	0.4979	−0.0021	0.0242	0.4990
	CEL	2.0619	0.0619	0.2515	2.0479	0.5009	0.0009	0.0375	0.4982
$n = 500, m = 1000$	PEL	2.0191	0.0191	0.1391	2.0042	0.5004	0.0004	0.0162	0.5009
	CEL	2.0250	0.0250	0.1594	2.0028	0.5007	0.0007	0.0239	0.4994
		$\pi = 0.2$							
$n = 200, m = 500$	PEL	1.9346	−0.0654	0.2134	1.9030	0.5342	0.0342	0.0234	0.5322
	CEL	1.9417	−0.0583	0.2279	1.9217	0.5343	0.0343	0.0370	0.5310
$n = 500, m = 1000$	PEL	1.8951	−0.1049	0.1190	1.8864	0.5371	0.0371	0.0166	0.5379
	CEL	1.9012	−0.0988	0.1385	1.8927	0.5374	0.0374	0.0250	0.5357
		$\pi = 0.5$							
$n = 200, m = 500$	PEL	1.6157	−0.3843	0.1285	1.5961	0.6380	0.1380	0.0225	0.6371
	CEL	1.6227	−0.3773	0.1467	1.6146	0.6382	0.1382	0.0411	0.6319
$n = 500, m = 1000$	PEL	1.6002	−0.3998	0.0714	1.5990	0.6402	0.1402	0.0166	0.6401
	CEL	1.6056	−0.3944	0.0892	1.6016	0.6400	0.1400	0.0273	0.6375

is obtained by fitting the external data with the Log-normal distribution. We set the true value $(\alpha, \beta) = (2, 0.5)$. The simulation is conducted based on $n = 200, m = 500$ and $n = 500, m = 1000$ over 200 replicates. The results are reported in Table 3. From this table, we can see that when π is close to the true one $\pi_0 = 0.1$, both PEL and CEL seem to be not too sensitive to the misspecified π values. However, when $\pi = 0.5$ is far away from the true value $\pi_0 = 0.1$, both PEL and CEL perform a little bit poorly. In general, the standard deviations of PEL estimators are still smaller than those of the CEL estimators.

For the robustness analysis about the misspecification of underlying incubation period density scenario, the internal data and external data are generated respectively from the mixture density function,

$$g(v, \boldsymbol{\eta}) = \pi f(v, \boldsymbol{\eta}) + (1 - \pi)\frac{\overline{F}(v, \boldsymbol{\eta})}{\mu},$$

with $\pi = 0, \ 0.05, \ 0.1, \ 0.2, \ 0.5$, where the true density function f of incubation period is chosen to be Gamma.

$$f(t) = \beta^\alpha t^{\alpha-1} \exp\{-(t\beta)\}/\Gamma(\alpha), \quad t > 0,$$

with $(\alpha, \beta) = (2, 0.5)$, but in our fitting, we treat it as the Weibull density. Again, the external auxiliary is derived by fitting the Log-normal density model. For $n = 200, m = 500$ and $n = 500, m = 1000$ over 200 replicates, the results are reported in

TABLE 4 Simulation results for sensitive analysis with the misspecification incubation period density function

	True	Estimation				
	$n = 200, m = 500$	$\pi = 0$	$\pi = 0.05$	$\pi = 0.1$	$\pi = 0.2$	$\pi = 0.5$
α	2.0000	1.3628	1.3795	1.3815	1.3998	1.4353
β	0.5000	0.2431	0.2405	0.2387	0.2356	0.2292
Mean	4.0000	3.8023	3.8340	3.8557	3.8944	3.9739
$Q_{0.05}$	0.7107	0.4583	0.4705	0.4798	0.4950	0.5396
$Q_{0.25}$	1.9226	1.6368	1.6646	1.6855	1.7204	1.8135
$Q_{0.50}$	3.3567	3.1462	3.1839	3.2111	3.2580	3.3705
$Q_{0.75}$	5.3853	5.2767	5.3185	5.3468	5.3978	5.4980
$Q_{0.90}$	7.7794	7.7136	7.7515	7.7738	7.8182	7.8695
$Q_{0.95}$	9.4877	9.3961	9.4273	9.4423	9.4770	9.4795
$Q_{0.975}$	11.1433	10.9850	11.0077	11.0139	11.0364	10.9837
$Q_{0.99}$	13.2767	12.9779	12.9871	12.9801	12.9838	12.8516
	$n = 500, m = 1000$	$\pi = 0$	$\pi = 0.05$	$\pi = 0.1$	$\pi = 0.2$	$\pi = 0.5$
α	2.0000	1.3422	1.3532	1.3621	1.3761	1.4193
β	0.5000	0.2418	0.2390	0.2381	0.2344	0.2287
Mean	4.0000	3.8460	3.8683	3.8800	3.9202	3.9883
$Q_{0.05}$	0.7107	0.4872	0.5010	0.5033	0.5212	0.5563
$Q_{0.25}$	1.9226	1.6828	1.7119	1.7188	1.7597	1.8360
$Q_{0.50}$	3.3567	3.1962	3.2313	3.2426	3.2956	3.3904
$Q_{0.75}$	5.3853	5.3237	5.3530	5.3690	5.4219	5.5095
$Q_{0.90}$	7.7794	7.7546	7.7629	7.7831	7.8201	7.8701
$Q_{0.95}$	9.4877	9.4332	9.4203	9.4430	9.4619	9.4726
$Q_{0.975}$	11.1433	11.0192	10.9821	11.0069	11.0043	10.9701
$Q_{0.99}$	13.2767	13.0100	12.9372	12.9644	12.9299	12.8303

Note: Table 4 summarizes the point estimates, mean, and quantiles, where α, β are the scale and shape parameters; Mean is the mean of incubation period, and $Q_{0.05}, Q_{0.25}, Q_{0.50}, Q_{0.75}, Q_{0.90}, Q_{0.95}, Q_{0.975}, Q_{0.99}$ denote the 5%, 25%, 50%, 75%, 90%, 95%, 97.5%, 99% quantiles of the incubation period. The parameters in the column corresponding to "True" are the theoretical parameters in the Gamma distribution; the parameters in the column corresponding to "Estimation" are the sample estimated parameters in the Weibull distribution.

Table 4. From this table, we can observe that our results are very robust and the underlying parameter estimates are close to the true ones.

4 | REAL DATA APPLICATION

The observed data are retrieved from municipal health committees in Zhejiang Province of China. Out of 1123 confirmed cases, the available information contains age, gender, 14 clinical symptoms such as fever and cough, date of symptom onset, and date of departure from Wuhan. From 19 January 2020 to 23 January 2020, there were 147 asymptomatic cases who left Wuhan and later developed symptoms in Zhejiang Province. Define the forward time as the elapse between departure from Wuhan and symptom onset. Note that before 19 January 2020, people were not aware of the severity of COVID-19. On the other hand, starting from 19 January 2020, the Center for Disease Control and Prevention of China began monitoring the outbreak of this epidemic. Various strict containment measures were implemented to minimize human-to-human transmission. Therefore, we are quite sure that all cases included in the analysis were infected in

TABLE 5 Analysis results for the incubation period of COVID-19

	PEL		CEL		MLE	
Parameters	Estimation	Confidence interval	Estimation	Confidence interval	Estimation	Confidence interval
α	2.36	(2.27, 2.53)	2.39	(1.90, 3.03)	2.58	(2.02, 3.95)
β	0.11	(0.11, 0.12)	0.12	(0.11, 0.15)	0.12	(0.11, 0.14)
Mean	7.75	(7.10, 7.99)	7.07	(5.97, 7.73)	7.22	(6.31, 8.16)
$Q_{0.05}$	2.48	(2.29, 2.61)	2.31	(1.42, 3.13)	2.57	(1.69, 4.01)
$Q_{0.25}$	5.16	(4.78, 5.31)	4.74	(3.52, 5.54)	5.01	(3.93, 6.34)
$Q_{0.50}$	7.48	(6.91, 7.71)	6.85	(5.59, 7.58)	7.05	(5.99, 8.12)
$Q_{0.75}$	10.04	(9.16, 10.37)	9.14	(7.96, 9.85)	9.22	(8.13, 10.18)
$Q_{0.90}$	12.45	(11.26, 12.95)	11.29	(10.08, 12.18)	11.23	(9.88, 12.33)
$Q_{0.95}$	13.92	(12.53, 14.53)	12.61	(11.13, 13.91)	12.44	(10.76, 13.74)
$Q_{0.975}$	15.19	(13.62, 15.91)	13.75	(12.01, 15.52)	13.48	(11.53, 15.03)
$Q_{0.99}$	16.69	(14.89, 17.56)	15.08	(13.05, 17.44)	14.69	(12.32, 16.67)

Note: Table 5 summarizes the point estimates and 95% confidence intervals of parameters, mean, and quantiles, where α, β are the parameters of the Weibull distribution; Mean is the sample mean of incubation period, and $Q_{0.05}, Q_{0.25}, Q_{0.50}, Q_{0.75}, Q_{0.90}, Q_{0.95}, Q_{0.975}, Q_{0.99}$ denote the 5%, 25%, 50%, 75%, 90%, 95%, 97.5%, 99% sample quantiles of the incubation period, respectively.

Wuhan. The last date of our collected data was 24 February 2020, which was also selected to ensure that those 147 cases met the criteria and assumptions in Qin et al,[1] that is, those cases had records of both dates of departure from Wuhan and dates of symptoms onset, and the follow-up times were long enough such that no additional biased sampling occurred in this study.

In our real data analysis, following Qin et al,[1] the density function of the forward time V, that is, the duration between departure from Wuhan and symptom onset in Zhejiang Province is

$$g(v) = \frac{\overline{F}(v)}{\mu}, \quad v \geq 0,$$

where $\overline{F}(\cdot)$ is the survival function of Weibull distribution with the density function

$$f(t) = \alpha\beta(t\beta)^{\alpha-1} \exp\{-(t\beta)^\alpha\}, \quad t > 0,$$

and $\mu = \int_0^\infty t f(t, \boldsymbol{\eta}) dt$ for $\boldsymbol{\eta} = (\alpha, \beta)^T$.

The auxiliary information is collected from Bi et al.[3] In Bi et al,[3] they treated the forward time as the time from arrival to symptom onset in Shenzhen for $m = 191$ travelers. By fitting those observed forward times with a Log-normal density function

$$h(v) = \frac{1}{v\sqrt{2\pi b}} \exp\left\{-\frac{1}{2}\left(\frac{\ln(v) - a}{b}\right)^2\right\},$$

they found the maximum likelihood estimators $\hat{a} = 1.22, \hat{b} = 0.85$. The covariance matrix of $(\hat{a}, \hat{b})$ can also be estimated by using the maximum likelihood estimation theory of misspecified models in White.[10]

Incorporating the individual-level forward time data with the auxiliary information from Bi et al,[3] we estimate the distribution of the incubation period using our proposed method. The results are reported in Table 5, where the 95% confidence intervals are obtained through 500 bootstrap replicates. The estimated mean and median of incubation period are 7.75 and 7.49 days, respectively. The 95%, 97.5%, and 99% percentiles are 13.92, 15.20, and 16.70 days, respectively. The maximum likelihood estimation (MLE), the classic empirical likelihood estimation (CEL) without auxiliary information, and their corresponding 95% confidence intervals are also obtained based on the observed data of Zhejiang Province. Our proposed PEL method produces shorter interval lengths compared with those derived from CEL method.

TABLE 6 Sensitivity Analysis results for the incubation period of COVID-19

Parameters	$\pi = 0.05$	$\pi = 0.1$	$\pi = 0.2$	$\pi = 0.5$
α	2.36 (2.21, 2.49)	2.25 (2.12, 2.46)	2.23 (1.91, 2.41)	1.79 (1.47, 2.11)
β	0.12 (0.11, 0.13)	0.12 (0.12, 0.13)	0.13 (0.12, 0.14)	0.16 (0.15, 0.17)
Mean	7.61 (6.92, 7.74)	7.26 (6.65, 7.39)	6.89 (6.31, 7.09)	5.64 (5.14, 5.94)
$Q_{0.05}$	2.44 (2.15, 2.53)	2.29 (1.98, 2.32)	2.05 (1.56, 2.21)	1.22 (0.76, 1.56)
$Q_{0.25}$	5.07 (4.58, 5.13)	4.81 (4.36, 4.82)	4.45 (3.81, 4.67)	3.17 (2.47, 3.57)
$Q_{0.50}$	7.36 (6.69, 7.44)	7.01 (6.44, 7.06)	6.61 (5.98, 6.81)	5.17 (4.45, 5.56)
$Q_{0.75}$	9.86 (8.96, 10.07)	9.43 (8.62, 9.67)	9.02 (8.28, 9.29)	7.61 (7.03, 7.94)
$Q_{0.90}$	12.23 (11.03, 12.58)	11.71 (10.67, 12.25)	11.32 (10.31, 11.74)	10.08 (9.28, 10.49)
$Q_{0.95}$	13.67 (12.30, 14.15)	13.11 (11.88, 13.88)	12.74 (11.51, 13.35)	11.67 (10.56, 12.40)
$Q_{0.975}$	14.93 (13.36, 15.52)	14.33 (12.93, 15.35)	13.99 (12.52, 14.79)	13.11 (11.67, 14.26)
$Q_{0.99}$	16.39 (14.63, 17.15)	15.75 (14.15, 17.01)	15.45 (13.72, 16.64)	14.83 (12.96, 16.57)

Table 6 summarizes the point estimates and 95% confidence intervals of parameters, mean, and quantiles, where α, β are the parameters of the Weibull distribution; Mean is the sample mean of incubation period, and $Q_{0.05}, Q_{0.25}, Q_{0.50}, Q_{0.75}, Q_{0.95}, Q_{0.975}, Q_{0.99}$ denote the 5%, 25%, 50%, 75%, 95%, 97.5%, 99% sample quantiles of the incubation period, respectively.

Next, we conduct a sensitivity analysis as in Qin et al[1] by fitting a mixture density function

$$g(v, \boldsymbol{\eta}) = \pi f(v, \boldsymbol{\eta}) + (1 - \pi)\frac{\overline{F}(v)}{\mu}, \tag{4}$$

where f and $\overline{F}(\cdot)/\mu$ are the density functions of incubation period and forward time, and π is the proportion that people contracted COVID-19 disease when they departed from Wuhan. In particular, when $\pi = 0$, the density function (4) simplifies to (1).

For $\pi = 0.05, 0.1, 0.2,$ and 0.5, the results are reported in Table 6. Compared with Table 5, the variations of the incubation period distribution estimation for $\pi = 0$ to 0.2 are small. However, the result for $\pi = 0.5$ is a little bit far away from the result for $\pi = 0$. We have already observed this behavior from Table 3 in our simulation study in last section. We believe it is unlikely that half of the COVID-19 patients were infected on their way out of Wuhan since in the early stage the number of infected patients is not big enough to infect others instantly from crowed environments, such as bus and train stations or airports. Qin et al[1] also did a similar sensitive analysis and found the range of π in $(0, 0.2]$ may be more reasonable.

5 | DISCUSSION

Some related works on utilizing empirical likelihood method to combine auxiliary information have been discussed, among others, by Qin,[7,12] Wu and Sitter,[13] Chaudhuri et al,[14] Rao and Wu,[15] Chatterjee et al,[16] Chaudhuri et al,[17] and Zhang et al.[18] In this article, we have proposed an efficient estimator to synthesize summarized information and individual level data. This approach can be viewed as a natural combination of the empirical likelihood method and confidence distribution.[19] It is shown that our method is theoretically valid and robust, even with the misspecified external model, and is more efficient compared with the traditional empirical likelihood method. Simulation studies also demonstrate that our method performs well under finite sample sizes and has a smaller variance compared with the classic empirical likelihood method without any external information. Moreover, similar to Owen's original empirical likelihood,[8] we have shown that the Wilks' theorem holds to be true for the augmented log-empirical likelihood.

The application of our method in estimating the incubation period of COVID-19 also provides important significance. By integrating the Zhejiang Province data and the external summarized information from Bi et al,[3] we find the estimated 95%, 97.5%, 99% quantiles are 13.92, 15.20, and 16.70 days, respectively. These numbers are smaller than those reported in Qin et al.[1] We believe this is due to small sample size problem since in general, a large sample size is needed in order to estimate the tail probability accurately. The sample sizes 147 and 191 from Zhejiang province and Shenzhen,

respectively, are much smaller than 1211 in Qin et al.[1] The existing reports, for example, Lauer et al[20] and Bi et al,[3] even from the same research team, may have contractive results due to small sample sizes and different data sets. Consider a simple case that we can observe incubation periods directly. If the true probability that the incubation period no less than 14 days is 0.1, then the 95% confidence interval error margin is $\pm 1.96 * \sqrt{0.9 * 0.1/147} = \pm 0.048$. This error margin would be even larger if only forward times are observed. Nevertheless, methods discussed in this article may shed light on combining auxiliary information in generalization of meta-analysis even if a misspecified model was used in the summarized information.

As a general guidance on the choice of π value from the range 0 to 0.2, we have following recommendation. If the histogram looks like monotonic in the entire range, then $\pi = 0$ is a good choice. Due to small sample size, from histogram plot we do not exclude the possibility that the underlying forward time density may not be strictly monotonic. Then $\pi = 0.2$ might be a good choice. Nevertheless, the differences in mean and quantiles approximately range a half day to one day.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study will be associated with our paper available. Once accepted, the data will be deposited to the repository Figshare, and will be linked to the final published article. Restrictions apply to the access to these data, which were used under agreement for this study.

ORCID
Baoying Yang 🄳 https://orcid.org/0000-0002-7198-7485

REFERENCES
1. Qin J, You C, Hu TJ, Yu SC, Zhou XH. Estimation of incubation period distribution of COVID-19 using disease onset forward time: a novel cross-sectional and forward follow-up study. *Sci Adv.* 2020;6.
2. Linton NM, Kobayashi T, Yang Y, et al. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *J Clin Med.* 2020;9(2):538.
3. Bi QF, Wu YS, Mei SJ, et al. Epidemiology and transmission of COVID-19 in Shenzhen China: analysis of 391 cases and 1286 of their close contacts. *Lancet Infect Dis.* 2020;20:911-919.
4. Lin DY, Zeng D. On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. *Biometrika.* 2010;97:321-332.
5. Liu DG, Liu RY, Xie MG. Multivariate meta-analysis of heterogeneous studies using only summary statistics: efficiency and robustness. *J Am Stat Assoc.* 2015;110:326-340.
6. Kundu P, Tang RL, Chatterjee N. Generalized meta-analysis for multiple regression models across studies with disparate covariate information. *Biometrika.* 2019;106:567-585.
7. Qin J. *Biased Sampling, Over-identified Parameter Problems and Beyond.* New York, NY: Springer; 2017.
8. Owen AB. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika.* 1988;75:237-249.
9. Cox DR. *Renewal Theory.* London, UK: Methuen; 1970.
10. White H. Maximum likelihood estimation of misspecified models. *Econometrica.* 1982;50:1-26.
11. DerSimonian R, Laird NM. Meta-analysis in clinical trials. *Contr Clin Trials.* 1986;7:177-188.
12. Qin J. Combining parametric and empirical likelihoods. *Biometrika.* 2000;87:484-490.
13. Wu C, Sitter RR. A model-calibration approach to using complete auxiliary information from survey data. *J Am Stat Assoc.* 2001;96:185-193.
14. Chaudhuri S, Handcock MS, Rendall MS. Generalised linear models incorporating population level information: an empirical likelihood based approach. *J R Stat Soc Ser B.* 2008;70:311-328.
15. Rao JNK, Wu C. Pseudo empirical likelihood inference for multiple frame surveys. *J Am Stat Assoc.* 2010;105:1494-1503.
16. Chatterjee N, Chen YH, Mass P, Carroll R. Constrained maximum likelihood estimation for model calibration using summary-level information from external big data sources. *J Am Stat Assoc.* 2016;111:107-117.
17. Chaudhuri S, Mondal D, Yin T. Hamiltonian Monte Carlo in Bayesian empirical likelihood computation. *J R Stat Soc Ser B.* 2017;79:293-320.
18. Zhang H, Deng L, Schiffman M, Qin J, Yu K. Generalized integration model for improved statistical inference by leveraging external summary data. *Biometrika.* 2020;107:689-703.
19. Singh K, Xie M, Strawderman W. Combining information from independent sources through confidence distributions. *Ann Stat.* 2005;33:159-183.

20. Lauer SA, Grantz KH, Bi QF, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Int Med.* 2020;172(9):577-582.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

APPENDIX A. REGULAR CONDITIONS

We assume the following conditions for theoretical derivation. For the convenience of description, denote $\alpha = \left(\eta^T, \theta^T\right)^T$, $\alpha_0 = \left(\eta_0^T, \theta_*^T\right)^T$, and $\boldsymbol{a}(v_i, \alpha) = \boldsymbol{a}(v_i, \eta, \theta)$.

C1:   The matrix $E\left[\partial \boldsymbol{a}^T(v_i, \alpha_0)/\partial\alpha\right]$ is of full rank and $E\left[\boldsymbol{a}(v_i, \alpha_0)\boldsymbol{a}(v_i, \alpha_0)^T\right]$ is positive definite.

C2:   The function $\partial \boldsymbol{a}^T(v, \alpha)/\partial\alpha$ and $\partial^2 \boldsymbol{a}_k(v, \alpha)/\partial\alpha\partial\alpha^T$ are continuous in a neighborhood of the true value $\left(\eta_0, \theta_*\right)^T$. In addition, $\|\partial \boldsymbol{a}^T(v, \alpha)/\partial\alpha\|$, $\|\partial^2 \boldsymbol{a}_k(v, \alpha)/\partial\alpha\partial\alpha^T\|$ and $\|\boldsymbol{a}(v, \alpha)\|^3$ are bounded by some integrable function $H(v)$ in that neighborhood, here $\boldsymbol{a}_k(v, \alpha)$ is the kth component of $\boldsymbol{a}(v, \alpha)$.

C3:   Assume that the external data and internal observations are independent and identically distributed.

The conditions C1 and C2 are commonly assumed in empirical likelihood literature, see, for example, Qin.[7] It is not difficult to verify that the first two conditions are met for

$$\boldsymbol{a}(\eta, \theta) = \begin{pmatrix} \frac{\partial \log h(v, \theta)}{\partial\theta} \\ \frac{\partial \log g(v, \eta)}{\partial\eta} \end{pmatrix}.$$

C3 is a mild condition and it holds in most practical situations.

APPENDIX B. PROOFS

Proof of the consistency and asymptotic normality

Let $\widetilde{\beta} = (\widetilde{\eta}^T, \widetilde{\theta}^T, \widetilde{\lambda}^T)^T$ be the solutions to the optimization problem (3), and denote $\beta = (\eta^T, \theta^T, \lambda^T)^T$ as the parameter vector. Following the asymptotic property of $\widehat{\theta}$ and the proof of theorem 8.2 in Qin,[7] we can easily see $(\widetilde{\eta}^T, \widetilde{\theta}^T, \widetilde{\lambda}^T)^T = (\eta_0^T, \theta_*^T, \boldsymbol{0}^T)^T + O_p(n^{-1/3})$. This completes the proof of consistency.

Now we derive the convergence rate. Firstly, we calculate the first derivative of ℓ_A.

$$\frac{\partial \ell_A}{\partial \eta} = -\sum_{i=1}^n \frac{1}{1 + \lambda^T \boldsymbol{a}(v_i, \eta, \theta)} \left(\frac{\partial \boldsymbol{a}^T(v_i, \eta, \theta)}{\partial \eta} \right) \lambda,$$

$$\frac{\partial \ell_A}{\partial \theta} = -\sum_{i=1}^n \frac{1}{1 + \lambda^T \boldsymbol{a}(v_i, \eta, \theta)} \left(\frac{\partial \boldsymbol{a}^T(v_i, \eta, \theta)}{\partial \theta} \right) \lambda - \widehat{\Sigma}^{-1}(\theta - \widehat{\theta}),$$

$$\frac{\partial \ell_A}{\partial \lambda} = -\sum_{i=1}^n \frac{\boldsymbol{a}(v_i, \eta, \theta)}{1 + \lambda^T \boldsymbol{a}(v_i, \eta, \theta)}.$$

By the Taylor expansion of $\partial \ell_A/\partial\beta$ around $(\eta_0^T, \theta_*^T, \boldsymbol{0}^T)^T$, we have

$$\boldsymbol{0} = \frac{1}{n} \frac{\partial \ell_A(\widetilde{\eta}, \widetilde{\theta}, \widetilde{\lambda})}{\partial \eta} = \frac{1}{n} \frac{\partial \ell_A(\eta_0, \theta_*, \boldsymbol{0})}{\partial \eta} + \frac{1}{n} \frac{\partial^2 \ell_A(\eta_0, \theta_*, \boldsymbol{0})}{\partial \eta\, \partial \lambda^T}(\widetilde{\lambda} - \boldsymbol{0}) + o_p(\delta_n),$$

$$\boldsymbol{0} = \frac{1}{n} \frac{\partial \ell_A(\widetilde{\eta}, \widetilde{\theta}, \widetilde{\lambda})}{\partial \theta} = \frac{1}{n} \frac{\partial \ell_A(\eta_0, \theta_*, \boldsymbol{0})}{\partial \theta} - \frac{1}{n}\widehat{\Sigma}^{-1}(\widetilde{\theta} - \theta_*) + \frac{1}{n} \frac{\partial^2 \ell_A(\eta_0, \theta_*, \boldsymbol{0})}{\partial \theta\, \partial \lambda^T}(\widetilde{\lambda} - \boldsymbol{0}) + o_p(\delta_n),$$

$$0 = \frac{1}{n}\frac{\partial \ell_A(\widetilde{\boldsymbol{\eta}}, \widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\lambda}})}{\partial \boldsymbol{\lambda}} = \frac{1}{n}\frac{\partial \ell_A(\boldsymbol{\eta}_0, \boldsymbol{\theta}_*, \mathbf{0})}{\partial \boldsymbol{\lambda}} + \frac{1}{n}\frac{\partial^2 \ell_A(\boldsymbol{\eta}_0, \boldsymbol{\theta}_*, \mathbf{0})}{\partial \boldsymbol{\lambda} \, \partial \boldsymbol{\eta}^T}(\widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$$

$$+ \frac{1}{n}\frac{\partial^2 \ell_A(\boldsymbol{\eta}_0, \boldsymbol{\theta}_*, \mathbf{0})}{\partial \boldsymbol{\lambda} \, \partial \boldsymbol{\theta}^T}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_*) + \frac{1}{n}\frac{\partial^2 \ell_A(\boldsymbol{\eta}_0, \boldsymbol{\theta}_*, \mathbf{0})}{\partial \boldsymbol{\lambda} \, \partial \boldsymbol{\lambda}^T}(\widetilde{\boldsymbol{\lambda}} - \mathbf{0}) + o_p(\delta_n),$$

where $\delta_n = \|\widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\| + \|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_*\| + \|\widetilde{\boldsymbol{\lambda}} - \mathbf{0}\|$. Thus, it holds that

$$\begin{pmatrix} \widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_0 \\ \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_* \\ \widetilde{\boldsymbol{\lambda}} - \mathbf{0} \end{pmatrix} = n \begin{pmatrix} \mathbf{0} & \mathbf{0} & \frac{\partial^2 \ell_A(\boldsymbol{\eta}_0,\boldsymbol{\theta}_*,\mathbf{0})}{\partial \boldsymbol{\eta}\,\partial \boldsymbol{\lambda}^T} \\ \mathbf{0} & -\widehat{\Sigma}^{-1} & \frac{\partial^2 \ell_A(\boldsymbol{\eta}_0,\boldsymbol{\theta}_*,\mathbf{0})}{\partial \boldsymbol{\theta}\,\partial \boldsymbol{\lambda}^T} \\ \frac{\partial^2 \ell_A(\boldsymbol{\eta}_0,\boldsymbol{\theta}_*,\mathbf{0})}{\partial \boldsymbol{\lambda}\,\partial \boldsymbol{\eta}^T} & \frac{\partial^2 \ell_A(\boldsymbol{\eta}_0,\boldsymbol{\theta}_*,\mathbf{0})}{\partial \boldsymbol{\lambda}\,\partial \boldsymbol{\theta}^T} & \frac{\partial^2 \ell_A(\boldsymbol{\eta}_0,\boldsymbol{\theta}_*,\mathbf{0})}{\partial \boldsymbol{\lambda}\,\partial \boldsymbol{\lambda}^T} \end{pmatrix}^{-1} \begin{pmatrix} o_p(\delta_n) \\ \frac{\widehat{\Sigma}^{-1}(\boldsymbol{\theta}_* - \widehat{\boldsymbol{\theta}})}{n} + o_p(\delta_n) \\ \frac{\sum_{i=1}^n \boldsymbol{a}(v_i, \boldsymbol{\eta}_0, \boldsymbol{\theta}_*)}{n} + o_p(\delta_n) \end{pmatrix}.$$

Since $\sum_{i=1}^n \boldsymbol{a}(v_i, \boldsymbol{\eta}_0, \boldsymbol{\theta}_*)/n = O_p(n^{-1/2})$ and $\widehat{\Sigma}^{-1}(\boldsymbol{\theta}_* - \widehat{\boldsymbol{\theta}})/n = O_p(n^{-1/2})$, it is clear that $\delta_n = O_p(n^{-1/2})$. That is, we have $(\widetilde{\boldsymbol{\eta}}^T, \widetilde{\boldsymbol{\theta}}^T, \widetilde{\boldsymbol{\lambda}}^T)^T = (\boldsymbol{\eta}_0^T, \boldsymbol{\theta}_*^T, \mathbf{0}^T)^T + O_p(n^{-1/2})$.

Next we show the asymptotic normality. From the Taylor expansion results of $\partial \ell_A/\partial \boldsymbol{\beta}$ and the law of large numbers, we have

$$\mathbf{0} = \frac{\partial \ell_A(\widetilde{\boldsymbol{\eta}}, \widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\lambda}})}{\partial \boldsymbol{\beta}}$$

$$= \frac{\partial \ell_A(\boldsymbol{\eta}_0, \boldsymbol{\theta}_*, \mathbf{0})}{\partial \boldsymbol{\beta}} + E\left[\frac{\partial^2 \ell_A(\boldsymbol{\eta}_0, \boldsymbol{\theta}_*, \mathbf{0})}{\partial \boldsymbol{\beta} \, \partial \boldsymbol{\beta}^T}\right]\begin{pmatrix} \widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_0 \\ \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_* \\ \widetilde{\boldsymbol{\lambda}} - \mathbf{0} \end{pmatrix} + o_p(n^{1/2}), \tag{B1}$$

where

$$E\left[\frac{\partial^2 \ell_A(\boldsymbol{\eta}_0, \boldsymbol{\theta}_*, \mathbf{0})}{\partial \boldsymbol{\beta} \, \partial \boldsymbol{\beta}^T}\right] = E \begin{pmatrix} \frac{\partial^2 \ell_A(\boldsymbol{\eta}_0,\boldsymbol{\theta}_*,\mathbf{0})}{\partial \boldsymbol{\eta}\,\partial \boldsymbol{\eta}^T} & \frac{\partial^2 \ell_A(\boldsymbol{\eta}_0,\boldsymbol{\theta}_*,\mathbf{0})}{\partial \boldsymbol{\eta}\,\partial \boldsymbol{\theta}^T} & \frac{\partial^2 \ell_A(\boldsymbol{\eta}_0,\boldsymbol{\theta}_*,\mathbf{0})}{\partial \boldsymbol{\eta}\,\partial \boldsymbol{\lambda}^T} \\ \frac{\partial^2 \ell_A(\boldsymbol{\eta}_0,\boldsymbol{\theta}_*,\mathbf{0})}{\partial \boldsymbol{\theta}\,\partial \boldsymbol{\eta}^T} & \frac{\partial^2 \ell_A(\boldsymbol{\eta}_0,\boldsymbol{\theta}_*,\mathbf{0})}{\partial \boldsymbol{\theta}\,\partial \boldsymbol{\theta}^T} & \frac{\partial^2 \ell_A(\boldsymbol{\eta}_0,\boldsymbol{\theta}_*,\mathbf{0})}{\partial \boldsymbol{\theta}\,\partial \boldsymbol{\lambda}^T} \\ \frac{\partial^2 \ell_A(\boldsymbol{\eta}_0,\boldsymbol{\theta}_*,\mathbf{0})}{\partial \boldsymbol{\lambda}\,\partial \boldsymbol{\eta}^T} & \frac{\partial^2 \ell_A(\boldsymbol{\eta}_0,\boldsymbol{\theta}_*,\mathbf{0})}{\partial \boldsymbol{\lambda}\,\partial \boldsymbol{\theta}^T} & \frac{\partial^2 \ell_A(\boldsymbol{\eta}_0,\boldsymbol{\theta}_*,\mathbf{0})}{\partial \boldsymbol{\lambda}\,\partial \boldsymbol{\lambda}^T} \end{pmatrix}$$

$$= -n E \begin{pmatrix} \mathbf{0} & \mathbf{0} & \frac{\partial \boldsymbol{a}^T(v_i,\boldsymbol{\eta},\boldsymbol{\theta})}{\partial \boldsymbol{\eta}} \\ \mathbf{0} & \widehat{\Sigma}^{-1}/n & \frac{\partial \boldsymbol{a}^T(v_i,\boldsymbol{\eta},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ \frac{\partial \boldsymbol{a}(v_i,\boldsymbol{\eta},\boldsymbol{\theta})}{\partial \boldsymbol{\eta}^T} & \frac{\partial \boldsymbol{a}(v_i,\boldsymbol{\eta},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} & -\boldsymbol{a}(v_i,\boldsymbol{\eta},\boldsymbol{\theta})\boldsymbol{a}^T(v_i,\boldsymbol{\eta},\boldsymbol{\theta}) \end{pmatrix}$$

$$\triangleq -n \begin{pmatrix} \mathbf{0} & \mathbf{0} & A_{13} \\ \mathbf{0} & A_{22} & A_{23} \\ A_{31} & A_{32} & -A_{33} \end{pmatrix}.$$

Here, the notation $\triangleq$ means 'definition', and obviously, $A_{ij} = A_{ji}^T$ for any $i, j = 1, \ldots, 3$. Hence

$$\sqrt{n}\begin{pmatrix} \widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_0 \\ \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_* \\ \widetilde{\boldsymbol{\lambda}} - \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & A_{13} \\ \mathbf{0} & A_{22} & A_{23} \\ A_{31} & A_{32} & -A_{33} \end{pmatrix}^{-1} \times \frac{1}{\sqrt{n}}\begin{pmatrix} \mathbf{0} \\ -\widehat{\Sigma}^{-1}(\boldsymbol{\theta}_* - \widehat{\boldsymbol{\theta}}) \\ -\sum_{i=1}^n \boldsymbol{a}(v_i, \boldsymbol{\eta}_0, \boldsymbol{\theta}_*) \end{pmatrix} + o_p(1). \tag{B2}$$

Furthermore, Equation (B2) implies that

$$
\sqrt{n}\begin{pmatrix} \widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_0 \\ \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_* \end{pmatrix} = J^{-1}UV^{-1}\frac{1}{\sqrt{n}}\begin{pmatrix} -\widehat{\Sigma}^{-1}(\boldsymbol{\theta}_* - \widehat{\boldsymbol{\theta}}) \\ -\sum_{i=1}^{n}\boldsymbol{a}(v_i, \boldsymbol{\eta}_0, \boldsymbol{\theta}_*) \end{pmatrix} + o_p(1),
$$

where

$$
U = \begin{pmatrix} 0 & A_{13} \\ A_{22} & A_{23} \end{pmatrix}, \quad V = \begin{pmatrix} A_{22} & 0 \\ 0 & A_{33} \end{pmatrix}, \quad J = UV^{-1}U^T,
$$

and according to Equation (B2), we have used the fact

$$
\begin{pmatrix} 0 & 0 \\ 0 & A_{22} \end{pmatrix}\begin{pmatrix} \widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_0 \\ \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_* \end{pmatrix} + \begin{pmatrix} A_{13} \\ A_{23} \end{pmatrix}\widetilde{\lambda} = \frac{1}{n}\begin{pmatrix} 0 \\ -\widehat{\Sigma}^{-1}(\boldsymbol{\theta}_* - \widehat{\boldsymbol{\theta}}) \end{pmatrix} + o_p\left(n^{-1/2}\right) \tag{B3}
$$

and

$$
\begin{pmatrix} A_{31} & A_{32} \end{pmatrix}\begin{pmatrix} \widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_0 \\ \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_* \end{pmatrix} - A_{33}\widetilde{\lambda} = -n^{-1}\sum_{i=1}^{n}\boldsymbol{a}(v_i, \boldsymbol{\eta}_0, \boldsymbol{\theta}_*) + o_p\left(n^{-1/2}\right).
$$

From this, easily we have

$$
\widetilde{\lambda} = A_{33}^{-1}(A_{31}, A_{32})\begin{pmatrix} \widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_0 \\ \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_* \end{pmatrix} + A_{33}^{-1}n^{-1}\sum_{i=1}^{n}\boldsymbol{a}(v_i, \boldsymbol{\eta}_0, \boldsymbol{\theta}_*) + o_p\left(n^{-1/2}\right). \tag{B4}
$$

Substituting (B4) into (B3), we have

$$
\begin{pmatrix} A_{13}A_{33}^{-1}A_{31} & A_{13}A_{33}^{-1}A_{32} \\ A_{23}A_{33}^{-1}A_{31} & A_{23}A_{33}^{-1}A_{32} + A_{22} \end{pmatrix}\begin{pmatrix} \widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_0 \\ \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_* \end{pmatrix}
$$
$$
= \begin{pmatrix} 0 & A_{13}A_{33}^{-1} \\ I & A_{23}A_{33}^{-1} \end{pmatrix} \times \frac{1}{n}\begin{pmatrix} -\widehat{\Sigma}^{-1}(\boldsymbol{\theta}_* - \widehat{\boldsymbol{\theta}}) \\ -\sum_{i=1}^{n}\boldsymbol{a}(v_i, \boldsymbol{\eta}_0, \boldsymbol{\theta}_*) \end{pmatrix} + o_p(n^{-1/2}).
$$

Then it is straightforward to verify that

$$
\sqrt{n}\begin{pmatrix} \widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_0 \\ \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_* \end{pmatrix} = J^{-1}UV^{-1}\frac{1}{\sqrt{n}}\begin{pmatrix} -\widehat{\Sigma}^{-1}(\boldsymbol{\theta}_* - \widehat{\boldsymbol{\theta}}) \\ -\sum_{i=1}^{n}\boldsymbol{a}(v_i, \boldsymbol{\eta}_0, \boldsymbol{\theta}_*) \end{pmatrix} + o_p(1) \tag{B5}
$$

with

$$
U = \begin{pmatrix} 0 & A_{13} \\ A_{22} & A_{23} \end{pmatrix}, \quad V = \begin{pmatrix} A_{22} & 0 \\ 0 & A_{33} \end{pmatrix}, \quad J = UV^{-1}U^T.
$$

Based on the asymptotic normality of $\widehat{\boldsymbol{\theta}}$, the independence between $n^{-1/2}\widehat{\Sigma}^{-1}(\boldsymbol{\theta}_* - \widehat{\boldsymbol{\theta}})$ and internal samples $\{v_i\}_{i=1}^{n}$, we can show

$$
\sqrt{n}\begin{pmatrix} \widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_0 \\ \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_* \end{pmatrix}
$$

asymptotically converges to a zero-mean normal distribution with covariance matrix

$$
\text{var}\left[\sqrt{n}\begin{pmatrix}\widetilde{\boldsymbol{\eta}}-\boldsymbol{\eta}_0\\\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_*\end{pmatrix}\right]=J^{-1}UV^{-1}\begin{pmatrix}A_{22}&0\\0&A_{33}\end{pmatrix}V^{-1}U^TJ^{-1}=J^{-1}
$$

$$
=\begin{pmatrix}A_{13}A_{33}^{-1}A_{31}&A_{13}A_{33}^{-1}A_{32}\\A_{23}A_{33}^{-1}A_{31}&A_{22}+A_{23}A_{33}^{-1}A_{32}\end{pmatrix}^{-1}. \tag{B6}
$$

This completes the proof.

Proof of Wilks' theorem

Denote $\widetilde{\boldsymbol{\beta}}_1=(\boldsymbol{\eta}_1^T,\widetilde{\boldsymbol{\theta}}_1^T,\widetilde{\boldsymbol{\lambda}}_1)^T$ as the maximum augmented log-empirical likelihood estimate under the null hypothesis H_0 : $\boldsymbol{\eta}_0=\boldsymbol{\eta}_1$, and let $\boldsymbol{\beta}_0=(\boldsymbol{\eta}_1^T,\boldsymbol{\theta}_*^T,\mathbf{0})^T$. From the proof of the consistency and asymptotic normality, we have $\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0=O_p(n^{-1/2})$. Applying the second-order Taylor expansion of $\ell_A(\boldsymbol{\eta},\boldsymbol{\theta},\lambda)$ for any $\boldsymbol{\beta}$ in the $O(n^{-1/2})$ neighborhood of the true value $\boldsymbol{\beta}_0$, we have

$$
\ell_A(\boldsymbol{\eta},\boldsymbol{\theta},\lambda)=\ell_A(\boldsymbol{\eta}_1,\boldsymbol{\theta}_*,0)+(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\frac{\partial\ell_A(\boldsymbol{\eta}_1,\boldsymbol{\theta}_*,0)}{\partial\boldsymbol{\beta}}
$$

$$
+\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^TE\left\{\frac{\partial^2\ell_A(\boldsymbol{\eta}_1,\boldsymbol{\theta}_*,0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right\}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+o_p(1).
$$

Similar to the result in (B4), if the derivative of $\partial\ell_A(\boldsymbol{\eta},\boldsymbol{\theta},\lambda)/\partial\lambda=0$, we can get

$$
\lambda=A_{33}^{-1}(A_{31},A_{32})\begin{pmatrix}\boldsymbol{\eta}-\boldsymbol{\eta}_1\\\boldsymbol{\theta}-\boldsymbol{\theta}_*\end{pmatrix}+A_{33}^{-1}n^{-1}\sum_{i=1}^n\boldsymbol{a}(v_i,\boldsymbol{\eta}_1,\boldsymbol{\theta}_*)+o_p\left(n^{-1/2}\right).
$$

After some algebra, we have

$$
\ell_A(\boldsymbol{\eta},\boldsymbol{\theta},\lambda)=\ell_A(\boldsymbol{\eta}_1,\boldsymbol{\theta}_*,0)-\frac{1}{2n}\sum_{i=1}^n\boldsymbol{a}^T(v_i,\boldsymbol{\eta}_0,\boldsymbol{\theta}_*)A_{33}^{-1}\sum_{i=1}^n\boldsymbol{a}(v_i,\boldsymbol{\eta}_0,\boldsymbol{\theta}_*)
$$

$$
+\begin{pmatrix}\boldsymbol{\eta}-\boldsymbol{\eta}_1\\\boldsymbol{\theta}-\boldsymbol{\theta}_*\end{pmatrix}^TUV^{-1}\begin{pmatrix}-\widehat{\Sigma}^{-1}(\boldsymbol{\theta}_*-\widehat{\boldsymbol{\theta}})\\-\sum_{i=1}^n\boldsymbol{a}(v_i,\boldsymbol{\eta}_0,\boldsymbol{\theta}_*)\end{pmatrix}
$$

$$
-\frac{n}{2}\begin{pmatrix}\boldsymbol{\eta}-\boldsymbol{\eta}_1\\\boldsymbol{\theta}-\boldsymbol{\theta}_*\end{pmatrix}^TJ\begin{pmatrix}\boldsymbol{\eta}-\boldsymbol{\eta}_1\\\boldsymbol{\theta}-\boldsymbol{\theta}_*\end{pmatrix}+o_p(1).
$$

Denote

$$
J=\begin{pmatrix}J_{11}&J_{12}\\J_{21}&J_{22}\end{pmatrix}.
$$

Differentiating ℓ_A with respect to $\boldsymbol{\theta}$ and using the fact that $\partial\ell_A\left(\widetilde{\boldsymbol{\eta}},\widetilde{\boldsymbol{\theta}},\widetilde{\lambda}\right)/\partial\boldsymbol{\beta}=0$, we have

$$
\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_*=-J_{22}^{-1}J_{21}\left(\widetilde{\boldsymbol{\eta}}-\boldsymbol{\eta}_1\right)-n^{-1}J_{22}^{-1}\left[\widehat{\Sigma}^{-1}(\boldsymbol{\theta}_*-\widehat{\boldsymbol{\theta}})+A_{23}A_{33}^{-1}\sum_{i=1}^n\boldsymbol{a}(v_i,\boldsymbol{\eta}_0,\boldsymbol{\theta}_*)\right]+o_p(n^{-1/2}).
$$

Let $\boldsymbol{\xi}_{1n}=-A_{13}A_{33}^{-1}\sum_{i=1}^n\boldsymbol{a}(v_i,\boldsymbol{\eta}_0,\boldsymbol{\theta}_*)$ and $\boldsymbol{\xi}_{2n}=-\widehat{\Sigma}^{-1}(\boldsymbol{\theta}_*-\widehat{\boldsymbol{\theta}})-A_{23}A_{33}^{-1}\sum_{i=1}^n\boldsymbol{a}(v_i,\boldsymbol{\eta}_0,\boldsymbol{\theta}_*)$, then

$$
\ell_A\left(\widetilde{\boldsymbol{\eta}},\widetilde{\boldsymbol{\theta}},\widetilde{\lambda}\right)=\ell_A(\boldsymbol{\eta}_1,\boldsymbol{\theta}_*,0)-\frac{1}{2n}\sum_{i=1}^n\boldsymbol{a}^T(v_i,\boldsymbol{\eta}_0,\boldsymbol{\theta}_*)A_{33}^{-1}\sum_{i=1}^n\boldsymbol{a}(v_i,\boldsymbol{\eta}_0,\boldsymbol{\theta}_*)+\frac{1}{2n}\boldsymbol{\xi}_{2n}^TJ_{22}^{-1}\boldsymbol{\xi}_{2n}^T
$$

$$
+(\widetilde{\boldsymbol{\eta}}-\boldsymbol{\eta}_1)^T(\boldsymbol{\xi}_{1n}-J_{12}J_{22}^{-1}\boldsymbol{\xi}_{2n})-\frac{n}{2}(\widetilde{\boldsymbol{\eta}}-\boldsymbol{\eta}_1)^T(J_{11}-J_{12}J_{22}^{-1}J_{21})(\widetilde{\boldsymbol{\eta}}-\boldsymbol{\eta}_1)^T+o_p(1).
$$

Similarly,

$$\ell_A\left(\boldsymbol{\eta}_1, \widetilde{\boldsymbol{\theta}}_1, \widetilde{\boldsymbol{\lambda}}_1\right) = \ell_A\left(\boldsymbol{\eta}_1, \boldsymbol{\theta}_*, 0\right) + \frac{1}{2n}\boldsymbol{\xi}_{2n}^T J_{22}^{-1}\boldsymbol{\xi}_{2n}^T$$

$$- \frac{1}{2n}\sum_{i=1}^{n}\boldsymbol{a}^T(v_i, \boldsymbol{\eta}_0, \boldsymbol{\theta}_*)A_{33}^{-1}\sum_{i=1}^{n}\boldsymbol{a}(v_i, \boldsymbol{\eta}_0, \boldsymbol{\theta}_*) + o_p(1).$$

Moreover, we have

$$R_E(\boldsymbol{\eta}_1) = 2\ell_A\left(\widetilde{\boldsymbol{\eta}}, \widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\lambda}}\right) - 2\ell_A\left(\boldsymbol{\eta}_1, \widetilde{\boldsymbol{\theta}}_1, \widetilde{\boldsymbol{\lambda}}_1\right)$$

$$= 2(\widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_1)^T(\boldsymbol{\xi}_{1n} - J_{12}J_{22}^{-1}\boldsymbol{\xi}_{2n}) - n(\widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_1)^T(J_{11} - J_{12}J_{22}^{-1}J_{21})(\widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_1)^T + o_p(1).$$

$$= \frac{1}{n}(\boldsymbol{\xi}_{1n} - J_{12}J_{22}^{-1}\boldsymbol{\xi}_{2n})^T(J_{11} - J_{12}J_{22}^{-1}J_{21})^{-1}(\boldsymbol{\xi}_{1n} - J_{12}J_{22}^{-1}\boldsymbol{\xi}_{2n}) + o_p(1).$$

where the last equation is from (B5). Finally, using the fact

$$\frac{1}{\sqrt{n}}\left(\boldsymbol{\xi}_{1n} - J_{12}J_{22}^{-1}\boldsymbol{\xi}_{2n}\right) = -\left(A_{13} - J_{12}J_{22}^{-1}A_{23}\right)A_{33}^{-1}\frac{\sum_{i=1}^{n}\boldsymbol{a}(v_i, \boldsymbol{\eta}_0, \boldsymbol{\theta}_*)}{\sqrt{n}} + J_{12}J_{22}^{-1}\frac{\widehat{\Sigma}^{-1}(\boldsymbol{\theta}_* - \widehat{\boldsymbol{\theta}})}{\sqrt{n}}$$

converges to a mean zero normal distribution with covariance matrix $J_{11} - J_{12}J_{22}^{-1}J_{21}$, we can conclude the proof of Wilks' theorem.