Supplement Material for "Assessment of Effects of Age and Gender on the Incubation Period of COVID-19 with a Mixture Regression Model"

SIMING ZHENG[1], JING QIN[2], AND YONG ZHOU*[3,4]

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

[2]National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, U.S.A.

[3]Key Laboratory of Advanced Theory and Application in Statistics and Data Science, MOE

[4]Academy of Statistics and Interdisciplinary Sciences, Faculty of Economics and Management, East China Normal University, Shanghai, China

1 Proof of Theorem 1

In order to prove the identification of regression coefficient $\boldsymbol{\beta}$, we first prove the following lemma:

Lemma 1. *If random variable V has density*

$$h^*(v;\alpha,\lambda,p) = \alpha\lambda\{p(\lambda v)^{\alpha-1} + (1-p)\Gamma\left(\frac{1}{\alpha}\right)\}\exp(-(\lambda v)^{\alpha}), v \geq 0 \tag{1}$$

with parameter $\alpha > 0, \lambda > 0, p \in [0,1]$, then all parameters in this density are identifiable for $\alpha \neq 1$, but scale parameter λ is always identifiable.

Proof of Lemma 1:

To prove scale parameter λ is identifiable, it suffices to show that for $(\alpha_1, \lambda_1, p_1)$ and $(\alpha_2, \lambda_2, p_2)$ with $\alpha_1, 2 > 0, \lambda_1, \lambda_2 > 0, p_1, p_2 \in [0,1]$, if

$$h^*(v;\alpha_1,\lambda_1,p_1) = h^*(v;\alpha_2,\lambda_2,p_2), \text{ for any } v \geq 0, \tag{2}$$

then $\lambda_1 = \lambda_2$.

Case 1($\alpha_1 = \alpha_2 = 1$): (2) is equivalent to

$$\lambda_1 \exp(-\lambda_1 v) = \lambda_2 \exp(-\lambda_2 v), \text{ for any } v \geq 0,$$

then it is obvious that $\lambda_1 = \lambda_2$ but p_1, p_2 can be arbitrary.

Case 2($\alpha_1 = 1, \alpha_2 \neq 1$): (2) is equivalent to

$$\lambda_1 \exp(-\lambda_1 v) = \alpha_2\lambda_2\{p_2(\lambda_2 v)^{\alpha_2-1} + (1-p_2)\Gamma\left(\frac{1}{\alpha_2}\right)\}\exp(-(\lambda_2 v)^{\alpha_2}), \text{ for any } v \geq 0,$$

then

$$\exp((\lambda_2 v)^{\alpha_2} - \lambda_1 v) = \alpha_2\lambda_2\{p_2(\lambda_2 v)^{\alpha_2-1} + (1-p_2)\Gamma\left(\frac{1}{\alpha_2}\right)\}/\lambda_1, \text{ for any } v \geq 0.$$

*Corresponding author. Email: yzhou@amss.ac.cn

Received April, 2020; Accepted May, 2020

Now as $\alpha_2 \neq 1$, the left term is the exponential form function of v but the right term is a polynomial function of v. Thus this is impossibly to happen.
Case 3($\alpha_1 \neq 1, \alpha_2 \neq 1$): Similarly, (2) is equivalent to

$$\exp((\lambda_2 v)^{\alpha_2} - (\lambda_1 v)^{\alpha_1}) = \frac{\alpha_2\lambda_2\{p_2(\lambda_2 v)^{\alpha_2-1} + (1-p_2)\Gamma\left(\frac{1}{\alpha_2}\right)\}}{\alpha_1\lambda_1\{p_1(\lambda_2 v)^{\alpha_1-1} + (1-p_1)\Gamma\left(\frac{1}{\alpha_1}\right)\}}, \text{ for any } v \geq 0.$$

Now let the left term be $m_1(v)$ and the right term $m_2(v)$. If $\alpha_1 \neq \alpha_2$, without loss of generality, suppose $\alpha_1 < \alpha_2$, then it must hold $\lim_{v\to\infty} \frac{m_1(v)}{m_2(v)} = \infty$, contradicting with $\frac{m_1(v)}{m_2(v)} \equiv 1$. Thus $\alpha_1 = \alpha_2 = a \neq 1$ for some constant a. Furthermore, if $\lambda_1 \neq \lambda_2$, again without loss of generality, suppose $\lambda_1 < \lambda_2$, then

$$\lim_{v\to\infty} m_1(v) = \lim_{v\to\infty} \exp(((\lambda_2)^a - (\lambda_1)^a)v^a) = \infty, \lim_{v\to\infty} m_2(v) = \frac{p_2\lambda_2^a}{p_1\lambda_1^a},$$

which contradicts with $\frac{m_1(v)}{m_2(v)} \equiv 1$. So it must hold $\lambda_1 = \lambda_2$. By the way, moreover, we know $p_1 = p_2$.

In short, if (2) holds, then $\lambda_1 = \lambda_2$. This completes the proof of lemma. □

Proof of Theorem 1:

To prove the identification of parameter of interest $\boldsymbol{\beta}$, it suffices to show that for $(\lambda_1, \alpha_1, \bar{\boldsymbol{\theta}}_1, \boldsymbol{\beta}_1)$ and $(\lambda_2, \alpha_2, \bar{\boldsymbol{\theta}}_2, \boldsymbol{\beta}_2)$, if

$$h(v|\boldsymbol{x}, \lambda_1, \alpha_1, \bar{\boldsymbol{\theta}}_1, \boldsymbol{\beta}_1) = h(v|\lambda_2, \alpha_2, \bar{\boldsymbol{\theta}}_2, \boldsymbol{\beta}_2), \forall\ v \geq 0, \boldsymbol{x} \in \mathcal{R}^p,$$

then $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. Now for any given $\boldsymbol{x}$, in the notation (1) of Lemma 1, we have

$$h^*(v; \alpha_1, \lambda_1 \exp(\boldsymbol{x}'\boldsymbol{\beta}_1), \pi(\boldsymbol{x}, \bar{\boldsymbol{\theta}}_1)) = h^*(v; \alpha_2, \lambda_2 \exp(\boldsymbol{x}'\boldsymbol{\beta}_2), \pi(\boldsymbol{x}, \bar{\boldsymbol{\theta}}_2)), \text{ for any } v \geq 0.$$

By the result of Lemma 1, we have $\lambda_1 \exp(\boldsymbol{x}'\boldsymbol{\beta}_1) = \lambda_2 \exp(\boldsymbol{x}'\boldsymbol{\beta}_2)$, for any $\boldsymbol{x} \in \mathcal{R}^p$ and therefore $\lambda_1 = \lambda_2, \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. What's more, by the analysis of Lemma 1, $\alpha_1 = \alpha_2$ and if $\alpha_1 = \alpha_2 \neq 1$, $\pi(\boldsymbol{x}, \bar{\boldsymbol{\theta}}_1) = \pi(\boldsymbol{x}, \bar{\boldsymbol{\theta}}_2)$ for any $\boldsymbol{x} \in \mathcal{R}^p$ and thus $\bar{\boldsymbol{\theta}}_1 = \bar{\boldsymbol{\theta}}_2$. This implies that in the case $\alpha \neq 1$ for the proposed mixture density, all parameters involved are identifiable. But parameter of interest $\boldsymbol{\beta}$ is always identifiable. This completes the proof of Theorem 1.□

2 Proof of Theorem 2

First, let $G(v|\boldsymbol{x}, \boldsymbol{\beta}, F, p) = pf(v\exp(\boldsymbol{x}'\boldsymbol{\beta}))\exp(\boldsymbol{x}'\boldsymbol{\beta}) + (1-p)g(v\exp(\boldsymbol{x}'\boldsymbol{\beta}))\exp(\boldsymbol{x}'\boldsymbol{\beta})$, where $f(v) = \frac{dF(v)}{dv}, g(v) = \frac{\bar{F}(v)}{\int \bar{F}(t)dt}, \bar{F}(t) = 1 - F_{\lambda,\alpha}(t)$. Suppose $G(v|\boldsymbol{x}, \boldsymbol{\beta}_0, F_0, \pi_0)$ is the true conditional density of V given $X = \boldsymbol{x}$ but we misspecify that it has the form $G(v|\boldsymbol{x}, \boldsymbol{\beta}, F_{\lambda,\alpha}, p)$ By the result of White (1982), we know that $\hat{\boldsymbol{\beta}}_{ML}$ is a consistent estimate of $\boldsymbol{\beta}^*$, which is a part of $(p^*, \lambda^*, \alpha^*, \boldsymbol{\beta}^*)$ and $(p^*, \lambda^*, \alpha^*, \boldsymbol{\beta}^*)$ maximizes the Kullback Leibler information

$$\int G(v|\boldsymbol{x}, \boldsymbol{\beta}_0, F_0, \pi_0) \log G(v|\boldsymbol{x}, \boldsymbol{\beta}, F, p)dv$$

Let $\bar{f}(v;\lambda,\alpha) = f_{\lambda,\alpha}(v), \bar{g}(v;\lambda,\alpha) = g_{\lambda,\alpha}(v)$ to emphasize the Weibull parameter, then as functions of v, we have

$$f_{\lambda,\alpha}(v\exp(\boldsymbol{x}'\boldsymbol{\beta}))\exp(\boldsymbol{x}'\boldsymbol{\beta}) = \bar{f}(v;\lambda\exp(\boldsymbol{x}'\boldsymbol{\beta}),\alpha), g_{\lambda,\alpha}(v\exp(\boldsymbol{x}'\boldsymbol{\beta}))\exp(\boldsymbol{x}'\boldsymbol{\beta}) = \bar{g}(v;\lambda\exp(\boldsymbol{x}'\boldsymbol{\beta}),\alpha)$$

$$\begin{aligned}
& \int G(v|x,\boldsymbol{\beta}_0,F_0,\pi_0)\log G(v|x,\boldsymbol{\beta},F_{\lambda,\alpha},p)dv \\
= & \int \left[\pi_0 f_0(v\exp(\boldsymbol{x}'\boldsymbol{\beta}_0))\exp(\boldsymbol{x}'\boldsymbol{\beta}_0) + (1-\pi_0)\frac{\bar{F}_0(v\exp(\boldsymbol{x}'\boldsymbol{\beta}_0))\exp(\boldsymbol{x}'\boldsymbol{\beta}_0)}{\int \bar{F}_0(t)dt}\right] \\
\cdot & \log\left[pf(v\exp(\boldsymbol{x}'\boldsymbol{\beta}))\exp(\boldsymbol{x}'\boldsymbol{\beta}) + (1-p)\frac{\exp(\boldsymbol{x}'\boldsymbol{\beta})\bar{F}(v\exp(\boldsymbol{x}'\boldsymbol{\beta}))}{\int \bar{F}(t)dt}\right]dv \\
= & \int \left[\pi_0 f_0(z) + (1-\pi_0)\frac{\bar{F}_0(z)}{\int \bar{F}_0(t)dt}\right] \\
\cdot & \log\left[pf(z\exp(\boldsymbol{x}'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)))\exp(\boldsymbol{x}'\boldsymbol{\beta}) + (1-p)\frac{\exp(\boldsymbol{x}'\boldsymbol{\beta})\bar{F}(z\exp(\boldsymbol{x}'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)))}{\int \bar{F}(t)dt}\right]dz \\
= & \int \left[\pi_0 f_0(z) + (1-\pi_0)\frac{\bar{F}_0(z)}{\int \bar{F}_0(t)dt}\right] \\
\cdot & \log\left[pf(z\exp(\boldsymbol{x}'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)))\exp(\boldsymbol{x}'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)) + (1-p)\frac{\exp(\boldsymbol{x}'(\boldsymbol{\beta}-\boldsymbol{\beta}_0))\bar{F}(z\exp(\boldsymbol{x}'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)))}{\int \bar{F}(t)dt}\right]dz + \boldsymbol{x}'\boldsymbol{\beta}_0 \\
= & \int \left[\pi_0 f_0(z) + (1-\pi_0)\frac{\bar{F}_0(z)}{\int \bar{F}_0(t)dt}\right]\left[p\bar{f}(z;\lambda\exp(\boldsymbol{x}'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)),\alpha) + (1-p)\bar{g}(z;\lambda\exp(\boldsymbol{x}'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)),\alpha)\right]dz \\
+ & \boldsymbol{x}'\boldsymbol{\beta}_0.
\end{aligned}$$

Let $g_0(z) = \pi_0 f_0(z) + (1-\pi_0)\frac{\bar{F}_0(z)}{\int \bar{F}_0(t)dt}, g(z;p,\sigma,\alpha) = p\bar{f}(z;\sigma,\alpha) + (1-p)\bar{g}(z;\sigma,\alpha)$, then we can find constants p_0, σ_0, α_0, such that

$$\int g_0(z)\log g(z;p,\sigma,\alpha)dz \le \int g_0(z)\log g(z;p_0,\sigma_0,\alpha_0)dz.$$

Thus

$$\begin{aligned}
& \int G(v|x,\beta_0,F_0,\pi_0)\log G(v|x,\beta,F,p)dv \\
= & \int g_0(z)\log g(z;p,\lambda\exp(\boldsymbol{x}'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)),\alpha)dz + \boldsymbol{x}'\boldsymbol{\beta}_0 \\
\le & \int g_0(z)\log g(z;p_0,\sigma_0,\alpha_0)dz + \boldsymbol{x}'\boldsymbol{\beta}_0.
\end{aligned}$$

And when $p = p_0, \lambda = \sigma_0, \boldsymbol{\beta} = \boldsymbol{\beta}_0, \alpha = \alpha_0$, the maximum is attained. Therefore, $(p^*, \lambda^*, \alpha^*, \boldsymbol{\beta}^*) = (p_0, \sigma_0, \alpha_0, \boldsymbol{\beta}_0)$ and $\hat{\boldsymbol{\beta}}_{ML}$ is a consistent estimate of $\boldsymbol{\beta}_0$. The proof is completed. □

3 EM Computation Algorithm

Let z be a binary random variable,

$$P(z = 1|\boldsymbol{x}) = \pi(\boldsymbol{x}, \boldsymbol{\theta}),$$

and

$$P(v,z|\boldsymbol{x}) = P(v|\boldsymbol{x},z)P(z|\boldsymbol{x}) = [\pi(\boldsymbol{x},\boldsymbol{\theta})f(v\exp(\boldsymbol{x}^T\boldsymbol{\beta}))\exp(\boldsymbol{x}^T\boldsymbol{\beta})]^z\left[(1-\pi(\boldsymbol{x},\boldsymbol{\theta}))\frac{\bar{F}(v\exp(\boldsymbol{x}^T\boldsymbol{\beta}))}{\int\bar{F}(t\exp(\boldsymbol{x}^T\boldsymbol{\beta}))dt}\right]^{1-z}.$$

The conditional likelihood is

$$\begin{aligned}
&\ln P(v,z|\boldsymbol{x})\\
=\;& z[\ln\pi(\boldsymbol{x},\boldsymbol{\theta}) + \ln f(v\exp(\boldsymbol{x}^T\boldsymbol{\beta}))\exp(\boldsymbol{x}^T\boldsymbol{\beta})] + (1-z)\left[\ln(1-\pi(\boldsymbol{x},\boldsymbol{\theta})) + \ln\frac{\bar{F}(v\exp(\boldsymbol{x}^T\boldsymbol{\beta}))}{\int\bar{F}(t\exp(\boldsymbol{x}^T\boldsymbol{\beta}))dt}\right]\\
=\;& z\ln\pi(\boldsymbol{x},\boldsymbol{\theta}) + (1-z)\ln(1-\pi(\boldsymbol{x},\boldsymbol{\theta})) + z\ln f(v\exp(\boldsymbol{x}^T\boldsymbol{\beta}))\exp(\boldsymbol{x}^T\boldsymbol{\beta}) + (1-z)\ln\frac{\bar{F}(v\exp(\boldsymbol{x}^T\boldsymbol{\beta}))}{\int\bar{F}(t\exp(\boldsymbol{x}^T\boldsymbol{\beta}))dt}
\end{aligned}$$

and the probability density for given $\boldsymbol{x}, z$ is

$$P(z=1|\boldsymbol{x},v) = \frac{P(v|\boldsymbol{x},z=1)P(z=1|\boldsymbol{x})}{P(v|\boldsymbol{x},z=1)P(z=1|\boldsymbol{x}) + P(v|\boldsymbol{x},z=0)P(z=0|\boldsymbol{x})}.$$

For the data $\{(v_i,\boldsymbol{x}_i)\}_{i=1}^n$, we present an EM algorithm as follow:

Step 1: Let $t=0$ and initialize $\hat{\lambda}_0, \hat{\alpha}_0, \hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\theta}}_0$ for the parameters $\lambda_I, \alpha_I, \boldsymbol{\beta}_I, \boldsymbol{\theta}_I$ and the tolerance $\epsilon > 0$

Step 2(E step): Let $t = t+1$,

$$\hat{\gamma}_k^t = P(z_k = 1|\boldsymbol{x}_k, v_k, \hat{\lambda}_{t-1}, \hat{\alpha}_{t-1}, \hat{\boldsymbol{\beta}}_{t-1}, \hat{\boldsymbol{\theta}}_{t-1}),$$

Given the observed data, the expectation of log full likelihood is

$$\begin{aligned}
L^t(\lambda,\alpha,\boldsymbol{\beta},\boldsymbol{\theta}) \;=\;& \sum_{k=1}^n\{\hat{\gamma}_k^t\ln\pi(\boldsymbol{x}_k,\boldsymbol{\theta}) + (1-\hat{\gamma}_k^t)\ln[1-\pi(\boldsymbol{x}_k,\boldsymbol{\theta})]\}\\
&+\sum_{k=1}^n\left\{\hat{\gamma}_k^t\ln[f(v_k\exp(\boldsymbol{x}_k^T\boldsymbol{\beta}))\exp(\boldsymbol{x}_k^T\boldsymbol{\beta})] + (1-\hat{\gamma}_k^t)\ln\frac{\bar{F}(v_k\exp(\boldsymbol{x}_k^T\boldsymbol{\beta}))}{\int\bar{F}(t\exp(\boldsymbol{x}_k^T\boldsymbol{\beta}))dt}\right\}.
\end{aligned}$$

Let

$$L_z^t(\boldsymbol{\theta}) = \sum_{k=1}^n\{\hat{\gamma}_k^t\ln\pi(\boldsymbol{x}_k,\boldsymbol{\theta}) + (1-\hat{\gamma}_k^t)\ln[1-\pi(\boldsymbol{x}_k,\boldsymbol{\theta})]\},$$

$$L_{xv}^t(\lambda,\alpha,\boldsymbol{\beta}) = \sum_{k=1}^n\left\{\hat{\gamma}_k^t\ln[f(v_k\exp(\boldsymbol{x}_k^T\boldsymbol{\beta}))\exp(\boldsymbol{x}_k^T\boldsymbol{\beta})] + (1-\hat{\gamma}_k^t)\ln\frac{\bar{F}(v_k\exp(\boldsymbol{x}_k^T\boldsymbol{\beta}))}{\int\bar{F}(t\exp(\boldsymbol{x}_k^T\boldsymbol{\beta}))dt}\right\}.$$

Step 3(M step):

$$\hat{\boldsymbol{\theta}}_t = \arg\max_{\boldsymbol{\theta}} L_z^t(\boldsymbol{\theta}), (\hat{\lambda}_t, \hat{\alpha}_t, \hat{\boldsymbol{\beta}}_t) = \arg\max_{\lambda,\alpha,\boldsymbol{\beta}} L_{xv}^t(\lambda,\alpha,\boldsymbol{\beta})$$

If $||(\hat{\lambda}_t, \hat{\alpha}_t, \hat{\boldsymbol{\beta}}_t', \hat{\boldsymbol{\theta}}_t')' - (\hat{\lambda}_{t-1}, \hat{\alpha}_{t-1}, \hat{\boldsymbol{\beta}}_{t-1}', \hat{\boldsymbol{\theta}}_{t-1}')'|| < \epsilon$, $\{\hat{\lambda}_t, \hat{\alpha}_t, \hat{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\theta}}_t\}$ is the final estimates. Otherwise, Return to Step 2.

On the convergence of EM algorithm, it may be refered to Wu (1983), Meng and Rubin (1993), Liu and Rubin (1994) and for more detailed literature review of EM algorithm, see the book by Liang et al. (2010).

References

Liang, F. M., C. H. Liu, and R. J. Carroll (2010). *Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples*. New York: Wiley.

Liu, C. H. and D. Rubin (1994). The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika 81*(4), 633–648.

Meng, X. L. and D. Rubin (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika 80*(2), 267–278.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica 50*(1), 1–26.

Wu, C. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics 11*, 95–103.