JIRSS (2025)
Vol. 24, No. 01, pp 67-96
DOI: 10.22034/jirss.2025.2046692.1086

Comparing Linear and non-linear Mixed Effects Models with Autoregressive Error under Small Area Estimation

Abdolreza Sayyareh[1], **Sedigheh Zamani Mehreyan**[2], **Omolbanin Bashiri Goudarzi**[3], **Mahmoud Torabi**[4]

[1] Department of Computer Sciences and Statistics, Faculty of Mathematics, K.N. Toosi, University of Technology, Tehran, Iran.
[2] Department of Statistics, Imam Khomeini International University, Qazvin, Iran.
[3] Department of Pure Mathematics, Faculty of Mathematics, K.N. Toosi, University of Technology, Tehran, Iran.
[4] Department of Biostatistics, Manitoba University, Winnipeg, Canada.

Received: 24 /11/2024, Accepted: 31/08/2025, Published online: 16/11/2025

Abstract. The problem of small area estimation is how to produce reliable estimates of characteristics of interest such as means, counts and quantiles. It is usually assumed that the observed values and the auxiliary values follow the linear regression model and the sampling errors are dependent and follow the autoregressive model. However, in practice, there are many situations in the observed values and the auxiliary values follow the non-linear regression model. We assume that the true model is unknown and consider some non-nested, non-linear or linear regression models as rival models and select an optimal model based on extensions of the model selection tests such as Vuong's test. This paper considers non-linear regression models to improve estimation and model selection based on latent variables and proposes a global model selection test for small-area estimation. A numerical example and real data analysis were carried out to illustrate the procedures obtained theoretically.

Keywords. EM Algorithm, Information Criterion, Model Selection, Mixed Effect Model, Small Area Estimation, Autoregressive Model.

MSC: 62p12, 62M10.

Corresponding Author: Abdolreza Sayyareh (asayyareh@kntu.ac.ir)
Sedigheh Zamani Mehreyan (s.zamani@sci.ikiu.ac.ir)
Omolbanin Bashiri Goudarzi (omolbaninbashiri68@gmail.com)
Mahmoud Torabi (mahmoud.torabi@umanitoba.ca).

1 Introduction

Small area estimation is becoming important in survey sampling due to growing demand for reliable small area statistics from both public and private sectors. The small sample sizes in the areas causes unacceptably large standard errors. Direct survey estimates for a small area, based on the data only from the sample units in the area, are likely to yield unacceptably large standard errors due to unduly small size of the sample in the area.

Let y_{iT} be the direct survey estimator of the ith small area mean for the current period T, $\mu_{iT}, i = 1, \cdots, M$, i.e.,

$$y_{iT} = \mu_{iT} + e_{iT},$$

where the e_{iT}'s are sampling errors with $E(e_{iT} \mid \mu_{iT}) = 0$, see Rao and Yu (1994). A simple regression synthetic estimator of the small area mean μ_{iT}, is obtained by the deterministic model on μ_{iT}:

$$\mu_{iT} = x_{iT}^{'}B,$$

where $\{(y_{iT}, x_{iT}), i = 1, \cdots, M\}$ is cross-sectional data for time T. Fay and Herriot (1979) proposed the model

$$\begin{aligned} y_{iT} &= \mu_{iT} + e_{iT}, \\ \mu_{iT} &= x_{iT}^{'}B + v_{iT}, \end{aligned} \tag{1.1}$$

where $x_{iT} = (x_{i1T}, \cdots, x_{ipT})^{'}$, $B = (B_1, \cdots, B_p)$ is the $p \times 1$ vector of regression coefficients, the v_i's are area-specific random effects assumed to be independent and identically distributed (iid) with mean 0 and unknown variance $Var(v_{iT}) = \sigma_v^2$, and the sampling errors e_{iT}'s are independent normal variables with

$$E(e_{iT}|\mu_{iT}) = 0, \;\; Var(e_{iT}|\mu_{iT}) = \sigma_{e_{iT}}^2,$$

where $\sigma_{e_{iT}}^2$ is known. Fay and Herriot (1979) have considered the empirical Bayes predictor, EB, for μ_i, Prasad and Rao (1990) and Lahiri and Rao (1995) have proposed the empirical best linear unbiased predictor, EBLUP, of μ_i. Jiang et al. (2011) derived the best predictive estimator of the fixed parameters under two well-known small area models, the Fay and Herriot model and the nested-error regression model.

The model (1.1) uses only cross- sectional data for the current period T. Many sample surveys are repeated in time with partial replacement of the sample elements. For such repeated surveys, considerable gain in efficiency can be achieved by borrowing strength across both areas and times. Since a design-based direct estimate of a small area characteristic based on the area sample is usually less reliable, model-based approach to small area estimation has become very popular to produce indirect estimates. Rao and Yu (1992, 1994) proposed an extension of the basic Fay and Herriot, FM, model as

$$\begin{aligned} y_{it} &= x_{it}^{'}B + v_i + u_{it} + e_{it}, \qquad i = 1, \cdots, M \;\; and \;\; t = 1, \cdots, T, \\ u_{it} &= \rho u_{i,t-1} + \epsilon_{it}, \qquad |\rho| < 1, \end{aligned} \tag{1.2}$$

to handle time series and cross-sectional data, where $\mathbf{y}_{it}$ is the value observed in area i at time t, $\mathbf{x}_{it} = (x_{i1t}, \cdots, x_{ipt})'$ is the covariate values associated with sample of i^{th} area at time t, the sampling errors $\mathbf{e} = (\mathbf{e_1}, \cdots, \mathbf{e_M})'$ are normally distributed with zero mean and block-diagonal covariance matrix Σ with arbitrary but known blocks $\sigma^2_{e_i}$. The $\sigma^2_{e_i}$ may be obtained by using the method of generalized variance function, see Wolter (1985), and stable estimate of $\sigma^2_{e_i}$, see Rao and Molina (2015). The random effects $\mathbf{v} = (v_1, \cdots, v_M)'$ are distributed with zero mean and variance $\sigma^2_v I_M$, $\epsilon = (\epsilon_1, \cdots, \epsilon_M)'$ are independent and normally distributed with $N(0, \sigma^2_\epsilon I_M)$, I_M is the identity matrix of order M, $\mathbf{u} = (u_1, \cdots, u_M)'$ are normally distributed with zero mean and covariance matrix $\sigma^2_\epsilon I_M \otimes \Gamma$, Γ is a $T \times T$ matrix with elements $\frac{\rho^{|i-j|}}{1-\rho^2}$ and $\otimes$ denotes Kronecker product. It is assumed that $\{v_i\}$ and $\{\epsilon_{it}\}$ are independent. In model-based small area estimation, it is usually assumed that there is a linear correlation between y_{it} and $\mathbf{x}_{it}$. In some cases, there is no linear correlation between y_{it} and $\mathbf{x}_{it}$ and this correlation is nonlinear, so a non-linear regression model can be suggested for the μ_i.

Modeling of regression parts is an important issue to obtain a good small area estimator, thereby variable selection problems would arise naturally, see Sugasawa and Kubokawa (2020). Vaida and Blanchard (2005) proposed the conditional Akaike Information Criterion, AIC, as an asymptotically unbiased estimator and pointed out that the classical AIC might be inappropriate for predicting quantity including random effects. Jiang et al. (2008) proposed fence method in generalized linear mixed models, and applied the method for selecting covariates in the FH model. Also, Jiang et al. (2010) adopted the fence method for fitting the P-spline models. This method requires selecting the degree of the spline, the number of knots, and a smoothing parameter. Torabi and Shokoohi (2014) considered hierarchical Bayes generalized linear models for a unified analysis of both discrete and continuous data with incorporating cross-sectional and time-series data. Torabi and Shokoohi (2015) proposed small area estimation under generalized linear mixed models using P-spline regression models to cover Normal and non-Normal responses and studied the empirical best predictor of small area parameters with corresponding prediction intervals. Shokoohi and Torabi (2018) proposed P-spline regression models for small-area estimation under the GLMMs. Also, they studied the empirical best predictors of small-area parameters and their corresponding prediction intervals where the maximum likelihood estimation approach is used to estimate the model parameters. Jiang et al. (2018) considered estimation of measure of uncertainty in small area estimation when a procedure of model selection is involved prior to the estimation. They proved the second-order unbiasedness of Monte-Carlo jackknife. Sugasawa et al. (2019) considered Fay and Herriot model, where the rival models are nested, and considered variable selection and estimation simultaneously to minimize the total mean squared prediction errors for estimation of small area means.

In this paper, we consider an extension of the Rao and Yu (1994) model and assume that the true model is unknown and propose non-linear regression models as rival models. The optimal model is selected based on the model selection test such as Vuong's test, Vuong (1989). We use latent variable to propose an algorithm which improves the estimators of unknown parameters. This test (unlike information criteria) has the ability to distinguish equivalent models. Sometimes Vuong's test selects two competing models as equivalent models. It is less clear that they are close to the unknown true model or far from it. Also, the EM algorithm has complex calculations and is affected by the initial values.The results of simulation show that this algorithm is dependent on the initial value for some complex models (especially for small sample sizes).

The rest of the paper is structured as follows: in Section 2, we consider linear and non-linear regression models and improve estimation and model selection based on latent variables. In Section 3, we study the obtained theoretical results by simulation. Also, we illustrate our theoretical results with the analysis of a real dataset in Section 4.

2 Estimation and Model Selection

In this section, we consider the problem of estimation and model selection in model-based small areas estimation. Let $y_{i1}, \cdots, y_{iT}$ for $i = 1, \cdots, M$ be the observed data for area i and suppose that y_{it}'s, $i = 1, \cdots, M$ and $t = 1, \cdots, T$, follow $f(y_{it}; \theta_{it})$ and

$$\begin{aligned} y_{it} &= \mu_{it} + e_{it}, \quad i = 1, \cdots, M \quad and \quad t = 1, \cdots, T, \\ \mu_{it} &= h_i(x_{it}, B) + v_i + u_{it}, \end{aligned} \tag{2.1}$$

where $h_i(x_{it}, B)$ denotes the non-linear regression model, the sampling errors e_{it} are independent and normally distributed with zero mean and variance $\sigma^2_{e_i}$ that it is known, $v_1, ..., v_M$ are i.i.d. $N(0, \sigma^2_v)$, which is independent of e_{it}'s. The u_{it}'s follow a common $AR(1)$ process for each i as

$$u_{it} = \rho u_{i,t-1} + \epsilon_{it}, \quad i = 1, \cdots, M \quad and \quad t = 1, \cdots, T,$$

where ϵ_{it}'s are independent and normally distributed with zero mean and variance σ^2_ϵ and $\{v_i\}$ and $\{\epsilon_{it}\}$ are independent. Model (2.1) can be rewritten as:

$$\begin{aligned} y_{it} &= h_i(x_{it}, B) + v_i + u_{it} + e_{it}, \quad i = 1, \cdots, M \quad and \quad t = 1, \cdots, T, \\ u_{it} &= \rho u_{i,t-1} + \epsilon_{it}, \mid \rho \mid < 1. \end{aligned} \tag{2.2}$$

By substituting $u_{it} = y_{it} - h_i(x_{it}, B) - v_i - e_{it}$ in model $u_{it} = \rho u_{i,t-1} + \epsilon_{it}$, we have

$$y_{it} - h_i(x_{it}, B) - v_i - e_{it} = \rho \left(y_{i,t-1} - h_i(x_{i,t-1}, B) - v_i - e_{i,t-1}\right) + \epsilon_{it}.$$

So, for $i = 1, \cdots, M$ and $t = 1, \cdots, T$,

$$\begin{aligned} y_{it} &= \rho y_{i,t-1} + h_i(x_{it}, B) - \rho h_i(x_{i,t-1}, B) + (1-\rho)v_i + e_{it} - \rho e_{i,t-1} + \epsilon_{it} \\ &= \rho y_{i,t-1} + h_i(x_{it}, B) - \rho h_i(x_{i,t-1}, B) + (1-\rho)v_i + \eta_{it} + \epsilon_{it}, \end{aligned}$$

where $\eta_{it} = e_{it} - \rho e_{i,t-1}$. The model (2.1) can be rewritten in matrix terms as

$$\mathbf{y}_t = \mathbf{h}(\mathbf{x}_t, \mathbf{B}) + \mathbf{v} + \mathbf{u}_t + \mathbf{e}_t, \quad t = 1, \cdots, T,$$

or equivalently

$$\mathbf{y}_t = \boldsymbol{\rho}\mathbf{y}_{t-1} + \mathbf{h}(\mathbf{x}_t, \mathbf{B}) - \boldsymbol{\rho}\mathbf{h}(\mathbf{x}_{t-1}, \mathbf{B}) + (1-\boldsymbol{\rho})I_M\mathbf{v} + \boldsymbol{\eta}_t + \boldsymbol{\epsilon}_t, \tag{2.3}$$

where $\mathbf{y}_t = (y_{1t}, \cdots, y_{Mt})'$, $\boldsymbol{\rho}$ is $M \times M$ matrix of unknown autoregressive coefficients, $\mathbf{h}(\mathbf{x}_t, \mathbf{B}) = (\mathbf{h}_1(\mathbf{x}_t, \mathbf{B}), \cdots, \mathbf{h}_M(\mathbf{x}_t, \mathbf{B}))'$, I_M is the $M \times M$ identity matrix, $\mathbf{v} = (v_1, \cdots, v_M)'$, $\boldsymbol{\eta}_t = (\eta_{1t}, \cdots, \eta_{Mt})'$ and $\boldsymbol{\epsilon}_t = (\epsilon_{it}, \cdots, \epsilon_{Mt})'$. In addition, $\mathbf{v}$, $\mathbf{e}_t$ and $\boldsymbol{\epsilon}_t$ are independently distributed with $\mathbf{v} \sim N_M(0, G)$, $G = \sigma_v^2 I_M$, $\boldsymbol{\epsilon}_t \sim N_M(0, S)$, $S = diag\left\{\sigma_{\epsilon_1}^2, \cdots, \sigma_{\epsilon_M}^2\right\} = diag\{\sigma_\epsilon^2\}$, $\mathbf{e}_t \sim N_M(0, \Sigma)$, $\Sigma = diag\left\{\sigma_{e_1}^2, \cdots, \sigma_{e_M}^2\right\}$ and $\boldsymbol{\eta} \sim N_M(0, \mathcal{S}), \mathcal{S} = \Sigma + \rho\Sigma\rho'$. Note that model (2.3) follows multivariate normal distribution with

$$\mu_y = (I_M - \boldsymbol{\rho})^{-1}\left(\mathbf{h}(\mathbf{x}_t, \mathbf{B}) - \boldsymbol{\rho}\mathbf{h}(\mathbf{x}_{t-1}, \mathbf{B})\right),$$

and variance - covariance matrix

$$V_y = G + \Sigma + (I_M - \boldsymbol{\rho}\boldsymbol{\rho}')^{-1}S.$$

The Expectation-Maximization, EM, algorithm is an efficient iterative procedure to compute the Maximum Likelihood, ML, estimate in the presence of missing or hidden data. The EM algorithm consists of two processes: In the Expectation-step, E-step, the missing data are estimated given the observed data and current estimate of the model parameters. In the Maximization-step, M-step, the likelihood function is maximized under the assumption that the missing data are known. In the following, based on the EM algorithm, an algorithm for improving the estimators in model-based small area estimation is presented.

Step 1: The obtained values of the maximum likelihood estimator of the model parameters are considered as start points.

Step 2: The hidden data are estimated given the observed data and current estimate of the model parameters.

Step 3: Under the assumption that the missing data are known, the likelihood function is maximized and the parameters of model (2.2) are computed.

2.1 EM Algorithm

Let $\mathbf{y}_c = (\mathbf{y}_t, \mathbf{v})'$ be complete data vector, where $\mathbf{y}_t$ be the observed data and $\mathbf{v}$ be the vector of unobserved or missing data. From model (2.3), we have that

$$\mathbf{y}_c = \begin{pmatrix} \mathbf{y}_t \\ \mathbf{v} \end{pmatrix} \sim N_{2M}\left(\begin{bmatrix} \boldsymbol{\mu}_y \\ 0 \end{bmatrix}, \begin{bmatrix} V_y & \sigma_v^2 I_M \\ \sigma_v^2 I_M & \sigma_v^2 I_M \end{bmatrix}\right),$$

where $\boldsymbol{\mu}_y = E(\mathbf{Y}_t)$, $\mathbf{V}_y = Var(\mathbf{Y}_t) = \mathbf{G} + \Sigma + \left(1 - \rho\rho'\right)^{-1} S$ and

$$E(Y_{it}) = (1-\rho)^{-1}\left(h_i(x_{it}, B) - \rho h_i(x_{i,t-1}, B)\right),$$

$$\begin{aligned} Cov(Y_{it}, v_i) &= Cov\left(\rho Y_{i,t-1} + h_i(x_{it}, B) - \rho h_i(x_{i,t-1}, B) + (1-\rho)v_i + \eta_{it} + \epsilon_{it}, v_i\right) \\ &= \rho Cov(Y_{i,t-1}, v_i) + (1-\rho)Cov(v_i, v_i) \\ &= \rho Cov(Y_{i,t-1}, v_i) + (1-\rho)\sigma_v^2. \end{aligned}$$

Based on the stationary condition, we have $Cov(Y_{it}, v_i) = \sigma_v^2$. Also,

$$Cov(Y_{it}, e_{it}) = Cov(\mu_{it} + e_{it}, e_{it}) = Cov(e_{it}, e_{it}) = \sigma_{e_i}^2,$$

and

$$\begin{aligned} Var(Y_{it}) &= \rho^2 Var(Y_{i,t-1}) + (1-\rho)^2 Var(v_i) + Var(\epsilon_{it}) + Var(\eta_{it}) \\ &+ 2\rho(1-\rho)Cov(Y_{i,t-1}, v_i) + 2\rho Cov(Y_{i,t-1}, \epsilon_{it}) + 2\rho Cov(Y_{i,t-1}, \eta_{it}) \\ &= \rho^2 Var(Y_{i,t-1}) + (1-\rho)^2\sigma_v^2 + \sigma_{\epsilon_i}^2 + (1+\rho^2)\sigma_{e_i}^2 \\ &+ 2\rho(1-\rho)\sigma_v^2 - 2\rho^2 Cov(Y_{i,t-1}, e_{i,t-1}). \end{aligned}$$

After some algebraic computation,

$$Var(Y_{it}) = \sigma_v^2 + \sigma_{e_i}^2 + \left(1 - \rho^2\right)\sigma_\epsilon^2.$$

Then, the distribution of $\mathbf{v}$ conditional on $\mathbf{Y}_t$ is

$$v|\mathbf{Y}_t = \mathbf{y}_t \sim N_M\left(\sigma_v^2 V_y^{-1}(\mathbf{Y}_t - \boldsymbol{\mu}_y), \sigma_v^2\left(I_M - \sigma_v^2 V_y^{-1}\right)\right).$$

The complete log- likelihood function can be expressed as

$$\begin{aligned} l(B, \sigma_v^2, \rho; y_c) &= l(B, \sigma_v^2, \rho; y|v) + l(\sigma_v^2, \rho; v) \\ &= -\frac{M}{2}\log(2\pi) - \frac{1}{2}\log\mid\varpi\mid \\ &- \frac{1}{2}\left(\left[\mathbf{y}_t - \boldsymbol{\rho}\mathbf{y}_{t-1} - \mathbf{h}(\mathbf{x}_t, \mathbf{B}) + \boldsymbol{\rho}\mathbf{h}(\mathbf{x}_{t-1}, \mathbf{B}) - (1-\boldsymbol{\rho})I_M\mathbf{v}\right]'\varpi^{-1}\right. \\ &\left.\times\left[\mathbf{y}_t - \boldsymbol{\rho}\mathbf{y}_{t-1} - \mathbf{h}(\mathbf{x}_t, \mathbf{B}) + \rho\mathbf{h}(\mathbf{x}_{t-1}, \mathbf{B}) - (1-\boldsymbol{\rho})I_M\mathbf{v}\right]\right) \\ &- \frac{M}{2}\log(2\pi) - \frac{1}{2}\log|\sigma_v^2 I_M| - \frac{1}{2\sigma_v^2}\mathbf{v}'\mathbf{v} \\ &= C - \frac{1}{2}\log\mid\varpi\mid - \frac{1}{2}\mathbf{z}_t'\varpi^{-1}\mathbf{z}_t - \frac{1}{2}\log|\sigma_v^2 I_M| - \frac{1}{2\sigma_v^2}\mathbf{v}'\mathbf{v}, \end{aligned}$$

where C is a constant that is independent of σ_v^2, $\varpi = S + \mathcal{S}$ and

$$\mathbf{z}_t = \mathbf{y}_t - \boldsymbol{\rho}\mathbf{y}_{t-1} - \mathbf{h}(\mathbf{x}_t, \mathbf{B}) + \boldsymbol{\rho}\mathbf{h}(\mathbf{x}_{t-1}, \mathbf{B}) - (1-\boldsymbol{\rho})I_M\mathbf{v}.$$

Let $\theta^{(0)} = \left(\mathbf{B}^{(0)}, \sigma_v^{2(0)}, \boldsymbol{\rho}^{(0)}\right)$ be starting values of $\theta = \left(\mathbf{B}, \sigma_v^2, \boldsymbol{\rho}\right)$, so

$$\begin{aligned} Q_1 &= E\left(l(B, \sigma_v^2, \rho; y_c)\right) \\ &= C - \frac{1}{2}\log\mid \varpi \mid - \frac{1}{2}E\left(\mathbf{z}_t'\boldsymbol{\varpi}^{-1}\mathbf{z}_t \mid \mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right) - \frac{1}{2}\log\mid \sigma_v^2 I_M \mid - \frac{1}{2\sigma_v^2}E\left(\mathbf{v}'\mathbf{v}|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right). \end{aligned} \quad (2.4)$$

The third term of Equation (2.4) is equal to

$$\begin{aligned} E\left(\mathbf{z}_t'\boldsymbol{\varpi}^{-1}\mathbf{z}_t|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right) &= E\left[tr\left(\mathbf{z}_t'\boldsymbol{\varpi}^{-1}\mathbf{z}_t|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right)\right] \\ &= E\left[tr\left(\boldsymbol{\varpi}^{-1}\mathbf{z}_t\mathbf{z}_t'|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right)\right] \\ &= tr\left[E\left(\boldsymbol{\varpi}^{-1}\mathbf{z}_t\mathbf{z}_t'|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right)\right] \\ &= tr\left[\boldsymbol{\varpi}^{-1}E\left(\mathbf{z}_t\mathbf{z}_t'|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right)\right] \\ &= tr\left[\boldsymbol{\varpi}^{-1}E\left(\mathbf{z}_t|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right)E\left(\mathbf{z}_t'|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right) + \boldsymbol{\varpi}^{-1}Var\left(\mathbf{z}_t|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right)\right] \\ &= E\left(\mathbf{z}_t'|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right)\boldsymbol{\varpi}^{-1}E\left(\mathbf{z}_t|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right) + tr\left[\boldsymbol{\varpi}^{-1}Var\left(\mathbf{z}_t|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right)\right], \end{aligned} \quad (2.5)$$

where

$$\begin{aligned} \tilde{\mathbf{z}}_1 &= E\left(\mathbf{z}_t|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right) \\ &= E\left[\mathbf{Y}_t - \boldsymbol{\rho}\mathbf{Y}_{t-1} - \mathbf{h}(\mathbf{x}_t, \mathbf{B}) + \boldsymbol{\rho}\mathbf{h}(\mathbf{x}_{t-1}, \mathbf{B}) - (1-\boldsymbol{\rho})\mathbf{v}|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right] \\ &= \mathbf{y}_t - \boldsymbol{\rho}^{(0)}\mathbf{y}_{t-1} - \mathbf{h}(\mathbf{x}_t, \mathbf{B}^{(0)}) + \boldsymbol{\rho}^{(0)}\mathbf{h}(\mathbf{x}_{t-1}, \mathbf{B}^{(0)}) - \left(1-\boldsymbol{\rho}^{(0)}\right)\tilde{b}_1, \end{aligned} \quad (2.6)$$

$$\tilde{b}_1 = E\left(\mathbf{v}|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right) = \sigma_v^{2(0)}V_y^{-1}(\mathbf{y}_t - \boldsymbol{\mu}_y^{(0)}),$$

and

$$\begin{aligned} Var\left(\mathbf{z}_t|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right) &= Var\left(\mathbf{Y}_t - \boldsymbol{\rho}\mathbf{Y}_{t-1} - \mathbf{h}(\mathbf{x}_t, \mathbf{B}) + \boldsymbol{\rho}\mathbf{h}(\mathbf{x}_{t-1}, \mathbf{B}) - (1-\boldsymbol{\rho})\mathbf{v}|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right) \\ &= (1-\boldsymbol{\rho}^{(0)})^2 Var\left(\mathbf{v}|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right) \\ &= \sigma_v^{2(0)}(1-\boldsymbol{\rho}^{(0)})^2\left(I_M - \boldsymbol{\sigma}_v^{2(0)}\mathbf{V}_y^{-1}\right). \end{aligned} \quad (2.7)$$

By substituting Equations (2.6) and (2.7) in Equation (2.5), we have

$$\begin{aligned} E\left(\mathbf{z}_t'\boldsymbol{\varpi}^{-1}\mathbf{z}_t|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right) &= \tilde{\mathbf{z}}_1'\boldsymbol{\varpi}^{-1}\tilde{\mathbf{z}}_1 + tr\left[\boldsymbol{\varpi}^{-1}Var\left(\mathbf{z}_t|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right)\right] \\ &= \tilde{\mathbf{z}}_1'\boldsymbol{\varpi}^{-1}\tilde{\mathbf{z}}_1 + tr\left[\boldsymbol{\varpi}^{-1}(1-\rho^{(0)})^2\sigma_v^{2(0)}(I_M - \sigma_v^{2(0)}V_y^{-1})\right]. \end{aligned} \quad (2.8)$$

The last term of Equation (2.4) is equal to

$$\begin{aligned} E(\mathbf{v}'\mathbf{v}|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}) &= E(\mathbf{v}'|\mathbf{y}_t, \boldsymbol{\theta}^{(0)})E(\mathbf{v}|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}) + tr\left[Var(\mathbf{v}|\mathbf{y}_t, \boldsymbol{\theta}^{(0)})\right] \\ &= \tilde{b}_1'\tilde{b}_1 + tr\left[\sigma_v^{2(0)}\left(I_M - \sigma_v^{2(0)}V_y^{-1}\right)\right]. \end{aligned} \quad (2.9)$$

By substituting Equations (2.8) and (2.9) in Equation (2.4), we have

$$\begin{aligned}
Q_1 &= E\left(l(B,\sigma_v^2,\rho;y_c)\right)\\
&= C - \frac{1}{2}\log|\varpi| - \frac{1}{2}\left(\tilde{\mathbf{z}}_1'\boldsymbol{\varpi}^{-1}\tilde{\mathbf{z}}_1 + tr\left[\boldsymbol{\varpi}^{-1}(1-\rho^{(0)})^2\sigma_v^{2(0)}(I_M - \sigma_v^{2(0)}V_y^{-1})\right]\right)\\
&- \frac{1}{2}\log|\sigma_v^2 I_M| - \frac{1}{2\sigma_v^2}\left(\tilde{b}_1'\tilde{b}_1 + tr\left[\sigma_v^{2(0)}\left(I_M - \sigma_v^{2(0)}V_y^{-1}\right)\right]\right).
\end{aligned} \tag{2.10}$$

We obtain estimators by solving the estimating equations as:

$$\begin{aligned}
0 = \frac{\partial Q_1}{\partial \mathbf{B}} &= \frac{\partial\left(\tilde{\mathbf{z}}_1'\boldsymbol{\varpi}^{-1}\tilde{\mathbf{z}}_1\right)}{\partial \mathbf{B}}\\
&= \frac{\partial}{\partial \mathbf{B}}\left\{\left(\mathbf{y}_t - \boldsymbol{\rho}^{(0)}\mathbf{y}_{t-1} - \mathbf{h}(\mathbf{x}_t,\mathbf{B}) + \boldsymbol{\rho}^{(0)}\mathbf{h}(\mathbf{x}_{t-1},\mathbf{B}) - (1-\boldsymbol{\rho}^{(0)})\tilde{b}_1\right)'\boldsymbol{\varpi}^{-1}\right.\\
&\times\left.\left(\mathbf{y}_t - \boldsymbol{\rho}^{(0)}\mathbf{y}_{t-1} - \mathbf{h}(\mathbf{x}_t,\mathbf{B}) + \boldsymbol{\rho}^{(0)}\mathbf{h}(\mathbf{x}_{t-1},\mathbf{B}) - (1-\boldsymbol{\rho}^{(0)})\tilde{b}_1\right)\right\},
\end{aligned} \tag{2.11}$$

$$0 = \frac{\partial Q_1}{\partial \boldsymbol{\rho}} = -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\rho}}\{\log|\varpi|\} - \frac{1}{2}\frac{\partial}{\partial \boldsymbol{\rho}}\left\{\tilde{\mathbf{z}}_1'\boldsymbol{\varpi}^{-1}\tilde{\mathbf{z}}_1 + (1-\rho)^2 tr(\boldsymbol{\varpi}^{-1}Var(\mathbf{v}|\mathbf{y}_t,\boldsymbol{\theta}^{(0)}))\right\}, \tag{2.12}$$

where $\varpi = S + \mathcal{S} = S + \Sigma + \rho\Sigma\rho'$ and

$$\tilde{z}_1 = \mathbf{y}_t - \boldsymbol{\rho}\mathbf{y}_{t-1} - \mathbf{h}(\mathbf{x}_t,\mathbf{B}^{(0)}) + \boldsymbol{\rho}\mathbf{h}(\mathbf{x}_{t-1},\mathbf{B}^{(0)}) - (1-\boldsymbol{\rho})\tilde{b}_1.$$

Also,

$$\hat{\sigma}_v^{2(1)} = \frac{1}{m}\left(\tilde{b}_1'\tilde{b}_1 + tr\left[Var(\mathbf{v}|\mathbf{y}_t,\boldsymbol{\theta}^{(0)}))\right]\right). \tag{2.13}$$

Thus, the EM algorithm is described as follows:

Step 0: Set $r = 0$ and choose starting value $\boldsymbol{\theta}^{(0)}$.

Step 1: For $r \geq 0$, calculate

$$\tilde{b}_1^{(r+1)} = E\left(\mathbf{v}|\mathbf{y}_t,\boldsymbol{\theta}^{(r)}\right) = \sigma_v^{2(r)}V_y^{-1}(\mathbf{y}_t - \boldsymbol{\mu}_y^{(r)}),$$

and

$$E(\mathbf{v}'\mathbf{v}|\mathbf{y}_t,\boldsymbol{\theta}^{(r)}) = \tilde{b}_1'\tilde{b}_1 + tr\left[\sigma_v^{2(r)}\left(I_M - \sigma_v^{2(r)}V_y^{-1}\right)\right].$$

Step 2: For $r \geq 0$, compute $\hat{\mathbf{B}}^{(r+1)}$, $\hat{\boldsymbol{\rho}}^{(r+1)}$ and $\hat{\sigma}_v^{2(r+1)}$ based on the Equations 2.11, 2.12 and 2.13.

Step 3. Iterate Steps 1 and 2 from $r = 1$ until reaching convergence. If the parameter update is smaller than a pre-specified threshold c, that is, if

$$\| \boldsymbol{\theta}^{(r)} - \boldsymbol{\theta}^{(r-1)} \| < c,$$

stop the algorithm, else return to step 1.

Although an EM iteration does increase the observed data likelihood function, no guarantee exists that the sequence converges to a maximum likelihood estimator. This means that an EM algorithm may converge to a local maximum of the observed data likelihood function, depending on starting values. A variety of heuristic or metaheuristic approaches exist to escape a local maximum, such as random-restart hill climbing (starting with several different random initial estimates) or applying simulated annealing methods. The EM algorithm is especially useful when the likelihood is an exponential family, see Sundberg (2019).

The tightness refers to how closely a given function describes the actual growth rate of another function as its input grows very large.

Let M denote the EM update, the result after one iteration of the EM algorithm. Recall from the general theory of the EM algorithm, $\hat{\theta}_n$ is a fixed point of M. Suppose that $\theta \to M(\theta)$ is differentiable and let $\lambda_n(\theta)$ be the largest eigenvalue of $D_\theta M(\theta) = Var(M(\theta))^{-1}$. Since the model is regular, then $0 \leq \lambda_n(\hat{\theta}_n) < 1$. Continuity (also a consequence of regularity) ensures that $\lambda_n(\theta)$ in a neighborhood of $\hat{\theta}_n$.

Proposition 1. Suppose that there is a $\lambda^* < 1$ such that $\lambda_n(\theta) \leq \lambda^*$ for all $\theta \in \Theta$ for n sufficiently large and almost every y-sequence. If there exists a $c < 1 - \lambda^*$ such that for some $C < \infty$,

$$E\left(\sqrt{n} \parallel \tilde{\theta}_n(1) - M(\tilde{\theta}_n(0)) \parallel_2 \mid \sqrt{n}(\tilde{\theta}_n(0) - \hat{\theta}_n) = h \right) \leq C + c \parallel h \parallel_2,$$

$\tilde{P}$-almost surely for n sufficiently (and almost every y-sequence) large, then the sequence $\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right)$ is tight.

Proof. See Nielsen (2000). □

Theorem 2.1. *Suppose $\sqrt{n}\left(\tilde{\theta}_n - \hat{\theta}_n\right)$ is tight for almost every y-sequence. Then, unconditionally*

$$\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right) \xrightarrow{D} N\left(0, I(\theta_0)^{-1}\left[2I - (I + F(\theta_0))^{-1}\right]\right),$$

where $\tilde{\theta}_n$, $\hat{\theta}_n$ and θ_0 denote the maximum likelihood estimator based on the EM algorithm, the maximum likelihood estimator and the true unknown value of θ, respectively. $I(\theta)$ denote the observed data information and $F(\theta)$ is the expected fraction of missing information.

Proof. It follows from Lemma 1 in Schenker and Welsh (1987), See Nielsen (2000). □

2.2 Vuong's Test for Model Selection

In this subsection, assume that $Y_{it}, t = 1, \cdots, T$ follow density function $f(y_{it}; \theta_{it})$ and

$$y_{it} = h_i(x_{it}, B) + v_i + u_{it} + e_{i,t}, \quad i = 1, \cdots, M \quad and \quad t = 1, \cdots, T,$$
$$u_{it} = \rho u_{i,t-1} + \epsilon_{it},$$

where the true model-based small area estimation is unknown. We consider two rival models-based small area estimation as:

$$y_{it} = \varrho y_{i,t-1} + \mathfrak{h}(x_{it}, \beta) - \varrho \mathfrak{h}(x_{i,t-1}, \beta) + (1-\varrho)v_i + \eta_{it} + \varepsilon_{it}, \quad \vartheta = (\beta, \varrho, \sigma_v^2) \subseteq R^p,$$

where $\mathfrak{h}_i(x_{it}, \beta)$ denote the non-linear regression model, $\eta_{it} = e_{it} - \rho e_{i,t-1}$, the sampling errors e_{it} are independent and normally distributed with zero mean and variance $\sigma_{e_i}^2$ that it is known, $v_1, \ldots, v_M$ are i.i.d. $N(0, \sigma_v^2)$, which is independent of e_{it}'s, ϱ is autoregressive coefficient, ε_{it}'s are independent and normally distributed with zero mean and variance σ_ε^2 and $\{v_i\}$ and $\{\varepsilon_{it}\}$ are independent, and

$$y_{it} = \phi y_{i,t-1} + k(x_{it}, \alpha) - \phi k(x_{it}, \alpha) + (1-\phi)V_i + \eta_{it} + \xi_{it}, \quad \nu = (\alpha, \phi, \sigma_V^2) \subseteq R^q,$$

where $k_i(x_{it}, \alpha)$ denote the non-linear regression model, the sampling errors e_{it} are independent and normally distributed with zero mean and variance $\sigma_{e_i}^2$ that it is known, $V_1, \ldots, V_M$ are i.i.d. $N(0, \sigma_V^2)$, which is independent of e_{it}'s, ϕ is autoregressive coefficient, ξ_{it}'s are normally distributed with zero mean and variance σ_ξ^2 and $\{V_i\}$ and $\{\xi_{it}\}$ are independent. Suppose non-linear models $h(x_{it}, \beta)$ and $k(x_{it}, \alpha)$ are non-nested models. Each model satisfies the Vuong's conditions, $A_1 - A_6$, that given in Appendix A.

Vuong (1989) proposed a Likelihood Ratio(LR)-based test for selecting between separate models. Define the complete-data likelihood ratio statistic for $f(y_{it}, \vartheta_{it})$ against $f(y_{it}, \nu_{it})$ as:

$$LR_c(\hat{\vartheta}_{it}, \hat{\nu}_{it}) = L_c(\hat{\vartheta}_{it}; y_{it}) - L_c^f(\hat{\nu}_{it}; y_{it}) = \sum_{t=1}^{T} \log\left(\frac{f(y_{it}; \hat{\vartheta}_{it})}{f(y_{it}; \hat{\nu}_{it})}\right).$$

Theorem 2.2. *(Asymptotic Distribution of the LR Statistic): Given assumptions A1-A6:*
(i) If $f(y_{it}; \vartheta_{it}^*) \neq f(y_{it}; \nu_{it}^*)$ then

$$T^{-\frac{1}{2}} LR_c(\hat{\vartheta}_{it}, \hat{\nu}_{it}) - T^{\frac{1}{2}} E_h\left[\log\left(\frac{f(y_{it}; \vartheta_{it}^*)}{f(y_{it}; \nu_{it}^*)}\right)\right] \xrightarrow{D} N(0, \omega_*^2),$$

where $V_{it}^* = c(\vartheta_{it}^*, \nu_{it}^*)$ are pseudo-true value of $V_{it} = c(\vartheta_{it}, \nu_{it})$, $\omega_*^2 = Var_h\left[\log\left(\frac{f(y_{it}; \vartheta_{it}^*)}{f(y_{it}; \nu_{it}^*)}\right)\right]$, E_h and Var_h denote the expectation and variance with respect to the true model.
(ii) If $f(y_{it}; \vartheta_{it}^*) = f(y_{it}; \nu_{it}^*)$ then

$$2LR_c(\hat{\vartheta}_{it}, \hat{\nu}_{it}) \xrightarrow{D} M_{p+q}(.; \lambda^*),$$

where M is weighted sum of chi-square distribution and λ^* is the vector of eigenvalues of $Q\Sigma_{V^*}$ and $\Sigma_{V^*} = I(V^*)^{-1}\left[2I - (I + F(V^*))^{-1}\right]$.

The proof of the theorem is given in Appendix B. Given a pair of rival models, it is natural to select the model that is close to the true model. It means that the model which has minimum Kullback-Leibler divergence, KL, is selected. Following Vuong (1989) we consider the following hypothesis:

$$H_0 : E_h \left\{ \log \left(\frac{f(y_{it}; \vartheta^*_{it})}{f(y_{it}; v^*_{it})} \right) \right\} = 0,$$

against $H_\vartheta : E_h \left\{ \log \left(\frac{f(y_{it}; \vartheta^*_{it})}{f(y_{it}; v^*_{it})} \right) \right\} > 0$ or $H_v : E_h \left\{ \log \left(\frac{f(y_{it}; \vartheta^*_{it})}{f(y_{i,t}; v^*_{it})} \right) \right\} < 0$. The null hypothesis mean that $f(y_{it}; \vartheta^*_{it})$ and $f(y_{it}; v^*_{it})$ are equivalent. The acceptance of H_ϑ means that $f(y_{it}; \vartheta^*_{it})$ is better than $f(y_{it}; v^*_{it})$ to fit the data. The acceptance of H_v means that $f(y_{it}; v^*_{it})$ is better than $f(y_{it}; \vartheta^*_{it})$ to fit the data.

Corollary 1. (Model Selection Test) Given assumptions A1-A6 and under H_0,

$$\frac{T^{-\frac{1}{2}} LR_c(\hat{\vartheta}_{it}, \hat{v}_{it})}{\omega_*} \xrightarrow{D} N(0,1).$$

Since $\frac{\omega_*}{\hat{\omega}_*} \xrightarrow{P} 1$, so using Slutsky Theorem $\frac{T^{-\frac{1}{2}} LR_c(\hat{\vartheta}_{it}, \hat{v}_{it})}{\hat{\omega}_*} \xrightarrow{D} N(0,1)$.

One chooses a critical value C_α from the standard normal distribution for some significance level. If $|\frac{T^{-\frac{1}{2}} LR_c(\hat{\vartheta}_{it}, \hat{v}_{it})}{\hat{\omega}_*}| < C_\alpha$ then one can not reject null hypothesis that the rival models are equivalent. If the value of the statistic $\frac{T^{-\frac{1}{2}} LR_c(\hat{\vartheta}_{it}, \hat{v}_{it})}{\hat{\omega}_*} > C_\alpha$, then one rejects the null hypothesis in favor of $f(y_{it}; v^*_{it})$ being better than $f(y_{it}; \vartheta^*_{it})$. If $\frac{T^{-\frac{1}{2}} LR_c(\hat{\vartheta}_{it}, \hat{v}_{it})}{\hat{\omega}_*}$ is smaller than $-C_\alpha$, then one rejects the null hypothesis in favor of $f(y_{it}; \vartheta^*_{it})$ being better than $f(y_{it}; v^*_{it})$.

Sayyareh (2012) studied the results of Vuong's test, Cox's test, Akaike's information criterion, Bayesian information criterion, Kullback information criterion and bias corrected Kullback information criterion and the ability of these tests to discriminate between non-nested linear models.

3 Simulation Study

In this Section we examine by simulation the relative performance of the material presented so far in the paper. In particular, we examine more closely the performance of the proposed algorithm for estimation and also the performance of the model selection test such as Vuong's test. The R software is used in the simulation and analysis of real data.

3.1 Part 1

Following simulation set up in Rao and Yu (1994) we have

$$\begin{aligned} y_{it} &= \rho y_{i,t-1} + h(x_{it}, B) - \rho h(x_{i,t-1}, B) + (1-\rho)v_i + e_{it} - \rho e_{i,t-1} + \epsilon_{it}, \\ &\mid \rho \mid < 1, \quad i = 1, \cdots, M \quad and \quad t = 1, \cdots, T, \end{aligned} \tag{3.1}$$

where $h(x_{i,t}, B) = B_0 \exp(B_1 x_{i,t})$, $e_{i,t}$'s independently distributed as $N(0, 1 + \frac{i}{T})$, v_i and $\epsilon_{i,t}$'s independently and identically distributed as $N(0,1)$, $\mu_\epsilon = 0$, $\sigma^2_\epsilon = 1$, $\rho = 0.7$, $B_0 = 2$ and $B_1 = 5$. The observations are generated from model (3.1), where $x_{i,t}$'s are generated from $N(0,2)$. We set $M = 30$ small areas and $T = 40$, and generate $R = 10^4$ independent samples $\{y^{(r)}_{it}; t = 1, \cdots, 40;\ i = 1, \cdots, m; r = 1, \cdots, R\}$. The samples are partitioned into two subsamples $\{y_{i,1}, \cdots, y_{i,32}\}$ and $\{y_{i,33}, \ldots, y_{i,40}\}$. The subsample $\{y_{i,1}, \cdots, y_{i,32}\}$ is used as training data and the subsample $\{y_{i,33}, \ldots, y_{i,40}\}$ is retained as testing data.

We consider the following rival models:

model 1: $y_{it} = \rho y_{i,t-1} + h(x_{it}, B) - \rho h(x_{i,t-1}, B) + (1-\rho) v_i + e_{it} - \rho e_{i,t-1} + \epsilon_{it}$,

$$h_1(x_{it}, B) = B_0 \exp(B_1 x_{it}),$$

model 2: $y_{it} = \rho y_{i,t-1} + h(x_{it}, B) - \rho h(x_{i,t-1}, B) + (1-\rho) v_i + e_{it} - \rho e_{i,t-1} + \epsilon_{it}$,

$$h_2(x_{it}, B) = B_0 (B_1 - x_{it})^{-1},$$

model 3: $y_{it} = \rho y_{i,t-1} + h(x_{it}, B) - \rho h(x_{i,t-1}, B) + (1-\rho) v_i + e_{it} - \rho e_{i,t-1} + \epsilon_{it}$,

$$h_3(x_{it}, B) = B_0 + B_1 \mathbf{x}_t.$$

It can be seen from Figure 1 that the relationship between y_{it} and x_{it} is non-linear, so we suggest model 2 . To compare the proposed model with the Rao and Yu (1992) model, we suggest model 3. Also to check the performance of the EM algorithm, we suggest model 1. In Table 17, we report the results of the mean of the values of estimators, mean of the values of the Bayesian information criterion, *BIC*, and mean-squared prediction error, $MSPE_h$,

$$MSPE_h = \frac{1}{8} \sum_{h=1}^{8} \left(y_{i,T+h} - \hat{y}_{i,T+h} \right)^2,$$

where $y_{i,T+h}$ is true value and $\hat{y}_{i,T+h}$ is h-step ahead forecast value. Recall that we estimate the parameters using the proposed EM algorithm. We do this to illustrate the potential problems that will arise in a misspecified model when parameters that do not belong to the true model are estimated and then used to make predictions. Since model 1 has the lowest value of the Bayesian information criterion, so this information criterion selects model 1 as the optimal model. Also, model 1 has the lowest value of the $MSPE_h$. The results of Table 1 show that rival models, model 2 and model 3, are not suitable for fitting the data based on the *BIC* and $MSPE_h$. Also, it show that the EM algorithm has a good performance in estimating model parameters.

Table 1: The value of estimated parameters of rival models and *BIC*.

	$\hat{\mu}_\epsilon$	$\hat{\sigma}^2_\epsilon$	$\hat{\rho}$	$\hat{B}_0$	$\hat{B}_1$	$MSPE_h$	BIC
model 1	-0.0013	1.0849	0.6531	2.0840	5.0416	1.0345	149.8534
model 2	3.2211	3.0827	0.5190	0.3155	2.3155	3.7182	227.0204
model 3	0.0000	2.2302	0.3032	3.2523	-1.6239	2.2647	200.2506

The estimated value of μ_v (mean of random effect v) and σ_v^2 (variance of random effect v) are computed and are presented in Table 2. It shows that although for all rival models, the estimated value of parameters are close to $\mu_v = 0$ and $\sigma_v^2 = 1$, but the estimated model 1 has the closest estimated value to the true value.

Table 2: The value of estimated parameters μ_v and σ_v^2.

model 1 $\hat{\mu}_v$	model 1 $\hat{\sigma}_v^2$	model 2 $\hat{\mu}_v$	model 2 $\hat{\sigma}_v^2$	model 3 $\hat{\mu}_v$	model 3 $\hat{\sigma}_v^2$
2.71018e-07	0.9998	7.1375e-06	0.9997	3.5195e-06	0.9995

For theoretical results of model selection test, using obtained data and proposed algorithm, the value of Vuong's statistic for paired of rival models are computed and compared with $z_{0.975}$. The relative frequency of proposed model selection test for each of rejection-acceptance regions is computed and the results are summarized in Tables 3, 4 and 5. We see that, based on the Vuong's test, the model 1 is better than the model 2. Also, it selects model 2 is better than the model 3. Table 3 presents the relative frequency of model selection for 30 areas when model 1 and model 2 are considered as rival models. The relative frequency of selecting model 1 as the optimal model (column 2) for each area is 1, meaning that model 1 is selected as the optimal model in every 10,000 iterations. The relative frequency of selecting model 2 as the optimal model (column 3) and selecting model 1 and model 2 as equivalent models (column 4) is 0, meaning that in none of the iterations is model 2 selected as the optimal model or model 1 and model 2 selected as equivalent models. The relative frequency of model selection for 30 areas is given in Table 4 when model 1 and model 3 are considered as rival models. The relative frequency of selecting model 1 as the optimal model (column 2) for each area is 1, meaning that model 1 is selected as the optimal model in every 10,000 iterations. In Table 5, the average of relative frequency of model selection is presented when model 2 and model 3 are considered as rival models. The average of relative frequency of selecting model 2 as the optimal model (column 2) is equal to 0.8924. The results of Tables 3 and 4 show that rival models, model 2 and model 3, are not suitable for fitting the data based on the Vuong's test. For more illustration see Figure 1. The graphs of the estimated model 1 and the splines estimated density function are almost the same, in other words, these two models are equivalent and also better than estimated models 2 and 3.

Table 3: The values of Vuong's test, $H_f : model\ 1 \quad VS \quad H_g : model\ 2$.

area	model 1 better	model 2 better	models 1 and 2 are equivalent
$(A_1, \ldots, A_{30})$	$(1, \ldots, 1)$	$(0, \ldots, 0)$	$(0, \ldots, 0)$

Table 4: The values of Vuong's test, $H_f : model\ 1 \quad VS \quad H_g : model\ 3$.

area	model 1 better	model 3 better	models 1 and 3 are equivalent
$(A_1, \ldots, A_{30})$	$(1, \ldots, 1)$	$(0, \ldots, 0)$	$(0, \ldots, 0)$

Table 5: The values of Vuong's test, H_f : *model* 2 VS H_g : *model* 3.

model 2 better	model 3 better	models 2 and 3 are equivalent
0.8924	0.0105	0.0971

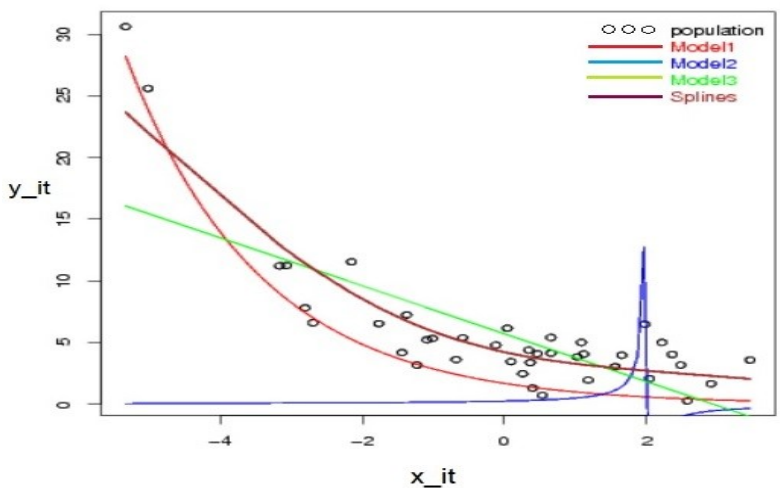

Figure 1: The generated data and estimated competing models curve.

To fit a proper model for data, we consider splines estimated density function as semi-parametric rival models. Semi-parametric estimation methods have some of the advantages over parametric estimation and some of the flexibility of non-parametric estimation. In a smooth spline, the time $t = 1, \ldots, T$ is divided into k intervals

$$[t_0, t_1],\ [t_1 + 1, t_2],\ \ldots,\ [t_{k-1} + 1, t_k = T],$$

with knots at $t_0, t_1, \ldots, t_k$. We then fit a polynomial regression of the form

$$f_t = B_0 + B_1 t + \cdots + B_p t^p,$$

in each interval. One of the most commonly used splines is cubic spline. For cubic spline, $p = 3$, the regression is fitted by minimizing

$$\sum_{t=1}^{T} (y_t - f_t)^2 + \lambda \int (f_t^{''})^2 dt,$$

where f_t is a cubic spline with a knot at each t. This optimization results in a compromise between the fit and the degree of smoothness, which is controlled by $\lambda \geq 0$.

Using data and estimated parameters, the values of Vuong's test for paired of parametric rival models and splines estimated density function (model 4) are computed. The results of the relative frequency of model selection test for each of rejection-acceptance regions are computed and the results are summarized in Tables 6-8. It shows if $a = 0.05$, the Vuong's test decides the splines density function and model 1 are equivalent. Also, it decides the splines estimated density function are better than model 2 and model 3.

Table 6: The values of Vuong's test, H_f : *model* 4 VS H_g : *model* 1.

area	model 4 better	model 1 better	models 1 and 4 are equivalent
$(A_1, \ldots, A_{30})$	$(0, \ldots, 0)$	$(0, \ldots, 0)$	$(1, \ldots, 1)$

Table 7: The values of Vuong's test, H_f : *model* 4 VS H_g : *model* 2.

area	model 4 better	model 2 better	models 2 and 4 are equivalent
$(A_1, \ldots, A_{30})$	$(1, \ldots, 1)$	$(0, \ldots, 0)$	$(0, \ldots, 0)$

Table 8: The values of Vuong's test, H_f : *model* 4 VS H_g : *model* 3.

area	model 4 better	model 3 better	models 3 and 4 are equivalent
$(A_1, \ldots, A_{30})$	$(1, \ldots, 1)$	$(0, \ldots, 0)$	$(0, \ldots, 0)$

3.2 Part 2

Similar to the simulation of part 1, the data is generated from model (3.1) where $h(x_{i,t}, B) = B_0 \exp(B_1 x_{i,t})$, $e_{i,t}$'s independently distributed as $N(0, 1 + \frac{i}{T})$, v_i and $\epsilon_{i,t}$'s independently and identically distributed as $N(0, 1)$, $\mu_\epsilon = 0$, $\sigma_\epsilon^2 = 1$, $\rho = 0.7, B_0 = 2$ and $B_1 = 0.5$. The observations are generated from model (3.1), where $x_{i,t}$'s are generated from $N_3(\mu_x, \Sigma_x)$,

$$\mu_x = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma_x = \begin{pmatrix} 1 & 2 & 0.4 \\ 2 & 5 & 0.5 \\ 0.4 & 0.5 & 3 \end{pmatrix}.$$

We set $M = 3$ small areas and $T = 30$, and generate $R = 10^4$ independent samples $\{y_{it}^{(r)}; t = 1, \cdots, 32;\ \ i = 1, \cdots, m; r = 1, \cdots, R\}$. Similar to the simulation of part 1, we consider $\{y_{i,1}, \cdots, y_{i,T}\}$ as training data, the subsample $\{y_{i,T+1}, y_{i,T+2}\}$ as testing data and model 1, model 2 and model 3 as rival models. The results of the mean of the values of estimators, mean of the values of the Bayesian information criterion, *BIC*, and mean-squared prediction error, $MSPE_h$, are given in Table 9. Since model 1 has the lowest value of the Bayesian information criterion, so this information criterion selects model 1 as the optimal model. Also, model 1 has the lowest value of the $MSPE_h$. The results of Table 9 show that rival models, model 2 and model 3, are not suitable for fitting the data based on the *BIC* and $MSPE_h$. Also, it show that the EM algorithm has a good performance in estimating model parameters.

Table 9: The value of estimated parameters of rival models and BIC.

Area	$\hat{\mu}_\epsilon$	$\hat{\sigma}^2_\epsilon$	$\hat{\rho}$	$\hat{B}_0$	$\hat{B}_1$	$MSPE_h$	BIC
model 1	-0.0155	1.0852	0.8404	2.0821	5.0257	1.2305	149.8486
model 2	2.7487	2.6333	0.5260	0.4803	0.4803	3.2964	210.9609
model 3	0.0000	1.9212	0.4177	2.9402	-1.4636	1.9801	184.7156

The estimated value of μ_v (mean of random effect v) and σ^2_v (variance of random effect v) are computed and are presented in Table 10. It shows that although for all rival models, the estimated value of parameters are close to $\mu_v = 0$ and $\sigma^2_v = 1$, but the estimated model 1 has the closest estimated value to the true value.

Table 10: The value of estimated parameters μ_v and σ^2_v.

model 1	model 1	model 2	model 2	model 3	model 3
$\hat{\mu}_v$	$\hat{\sigma}^2_v$	$\hat{\mu}_v$	$\hat{\sigma}^2_v$	$\hat{\mu}_v$	$\hat{\sigma}^2_v$
-0.0155	1.0852	2.7487	2.6333	-1.1892e-09	1.9212

The value of Vuong's statistic for paired of rival models are computed and compared with $z_{0.975}$. The relative frequency of proposed model selection test for each of rejection-acceptance regions is computed and the results are summarized in Tables 11, 12 and 13. They show that, based on the Vuong's test, the model 1 is better than the model 2 and model 3. Also, it selects model 2 and model 3 as equivalent models.

Table 11: The values of Vuong's test, $H_f : model\ 1 \quad VS \quad H_g : model\ 2$.

area	model 1 better	model 2 better	models 1 and 2 are equivalent
$(A_1, \dots, A_2, A_3)$	$(0.9839, 1, 1)$	$(0, 0, 0)$	$(0.0161, 0, 0)$

Table 12: The values of Vuong's test, $H_f : model\ 1 \quad VS \quad H_g : model\ 3$.

area	model 1 better	model 3 better	models 1 and 3 are equivalent
(A_1, A_2, A_3)	$(0.9654, 0.9991, 0.9873)$	$(0, 0, 0)$	$(0.0346, 0.0009, 0.0127)$

Table 13: The values of Vuong's test, $H_f : model\ 2 \quad VS \quad H_g : model\ 3$.

model 2 better	model 3 better	models 2 and 3 are equivalent
(0.0211, 0.0107, 0.0033)	(0,0,0)	(0.9789, 0.9893, 0.9967)

The values of Vuong's test for paired of parametric rival models and splines estimated density function (model 4) are computed. The results of the relative frequency of model selection test for each of rejection-acceptance regions are computed and the results are summarized in Tables 14-16. It shows if a = 0.05, the Vuong's test decides the splines density function and model 1 are equivalent. Also, it decides the splines estimated density function are better than model 2 and model 3.

From the results of both simulations, it can be seen that although the obtained results of the first simulation are better, the difference in these results is not significant and both simulations confirm the same result. Both simulations show that model 1 and the semi-parametric model are equivalent and are better than other rival models.

Table 14: The values of Vuong's test, H_f : *model* 4 VS H_g : *model* 1 .

area	model 4 better	model 1 better	models 1 and 4 are equivalent
(A_1, A_2, A_3)	$(0,0,0)$	$(0,0,0)$	$(1,1,1)$

Table 15: The values of Vuong's test, H_f : *model* 4 VS H_g : *model*2.

area	model 4 better	model 2 better	models 2 and 4 are equivalent
(A_1, A_2, A_3)	$(1,1,1)$	$(0,0,0)$	$(0,0,0)$

Table 16: The values of Vuong's test, H_f : *model* 4 VS H_g : *model* 3.

area	model 4 better	model 3 better	models 3 and 4 are equivalent
(A_1, A_2, A_3)	$(1,1,1)$	$(0,0,0)$	$(0,0,0)$

4 Real Data Analysis

Climate data analysis is the most fundamental step in predicting the c1imate change. lts main objective is to increase understanding of the atmosphere and its interaction with the oceans, cry sphere and the land surface, through various types of approaches or techniques. The humans observe nature and climate from antique time to now to find out, or take out information, and to predict. The climate system is complex because not all variables can be observed at random spatial and temporal range and resolution, our information and shall be controlled and improbability is introduced. To handle such complexity in the climate system, climate data analysis tools and software plays a vital role to obtain quantitative and qualitative information to predict the weather.

Climate is the result of a slight constancy between several elements like atmosphere, water systems, living organisms and topography. These elements establish various factors that manage the climate. These factors are called climatic variables. The most important and common climate variables are rain, atmospheric pressure, wind, humidity and temperature. These climate variables are essential in climate data analysis, research.

We considered the climate change Canada dataset from the website of Environment and Climate Change Canada [1]. This web site provides basic access to selected data and locations contained in Environment and Climate Change Canada's database of

[1] *https : //climate.weather.gc.ca/historical_data/search_historic_data_e.html.*

historical climate data. This data corresponds to three basic sampling frequencies of climate data collection:

- Hourly data is provided for each hour of the day requested
- Daily data is provided for each day of the month requested
- Monthly data is provided for each month of the year requested

Data collection, processing, quality control checks and procedures have evolved and changed over the years since the earliest data was observed in 1840. Data contains information about monthly, seasonal and annual means of the daily maximum, minimum and mean temperatures, daily rainfall, snowfall and total precipitation for many locations in Canada. There are thirteen provinces. These thirteen provinces were further sub-divided into 8819 stations, with data available between 1840 and 2023. We use a monthly returns of climate change Canada from January 2022 to December 2022. The 1545 stations found within all provinces and territories, with data available between January 2022 and December 2022. The data of some stations were not complete. Since the purpose was to evaluate the performance of the EM algorithm and the model selection test for small areas, we considered four provinces (British Columbia, Ontario, New Brunswick, and Alberta) and selected 20 stations from these provinces with complete data, as listed in Table 17.

Table 17: Twenty selected weather stations from four Canadian provinces (British Columbia, Ontario, New Brunswick, Alberta) with complete monthly data for 2022, used for evaluating the EM algorithm and model selection.

Parameter	Station	Province
A_1	100 MILE HOUSE 6NE	BRITISH COLUMBIA
A_2	AMHERSTBURG	ONTARIO
A_3	APPLETON	ONTARIO
A_4	AROOSTOOK	NEW BRUNSWICK
A_5	KAMLOOPS PRATT ROAD	BRITISH COLUMBIA
A_6	BALDWIN	ONTARIO
A_7	BARKERVILLE	BRITISH COLUMBIA
A_8	BELLEVILLE	ONTARIO
A_9	CAPE BEALE LIGHT	BRITISH COLUMBIA
A_{10}	KINGSVILLE MOE	ONTARIO
A_{11}	CENTREVILLE	ONTARIO
A_{12}	CHEMAINUS	BRITISH COLUMBIA
A_{13}	CORNER BROOK	NEWFOUNDLAND
A_{14}	DELTA TSAWWASSEN BEACH	BRITISH COLUMBIA
A_{15}	KOOTENAY NP WEST GATE	BRITISH COLUMBIA
A_{16}	ENTWISTLE	ALBERTA
A_{17}	FABYAN	ALBERTA
A_{18}	LITTLE QUALICUM HATCHERY	BRITISH COLUMBIA
A_{19}	GRIMSBY MOUNTAIN	ONTARIO
A_{20}	HALIBURTON 3	ONTARIO

These datasets contain the monthly returns with the sample extending from 01-Jan-2021 to the 31-Dec- 2021. We used the mean temperatures as dependent variable and daily snowfall as covariate variable and denote $r_{t,i}$ as the i^{th} daily return for the t^{th} month then the monthly data is defined as $\sum_{i=1}^{m} r_{i,t}$, where m is the number of days. Consider model,

$$y_{it} = \rho y_{i,t-1} + h(x_{it}, B) - \rho h(x_{i,t-1}, B) + (1-\rho)v_i + \beta_1 cos(2\pi f t) + \beta_2 sin(2\pi f t) + \eta_{it} + \epsilon_{it}, \qquad (4.1)$$

as family of rival models, where $f = \frac{1}{12}$ and $\beta_1 cos(2\pi f t) + \beta_2 sin(2\pi f t) + \eta_{it}$ capture seasonality of the data. For each area, we consider four rival models:

model 5:

$$y_{it} = \rho y_{i,t-1} + h_5(x_{it}, B) - \rho h_5(x_{i,t-1}, B) + (1-\rho)v_i + \beta_1 cos(2\pi f t) + \beta_2 sin(2\pi f t) + \eta_{it} + \epsilon_{it},$$
$$h_5(x_{it}, B) = B_0 \exp(-B_1 x_{it})) - B_3 x_{it},$$

model 6:

$$y_{it} = \rho y_{i,t-1} + h_6(x_{it}, B) - \rho h_6(x_{i,t-1}, B) + (1-\rho)v_i + \beta_1 cos(2\pi f t) + \beta_2 sin(2\pi f t) + \eta_{it} + \epsilon_{it},$$
$$h_6(x_{it}, B) = B_0(1 + B_1 x_{it})^{-1} - B_3 x_{it},$$

model 7:

$$y_{it} = \rho y_{i,t-1} + h_7(x_{it}, B) - \rho h_7(x_{i,t-1}, B) + (1-\rho)v_i + \beta_1 cos(2\pi f t) + \beta_2 sin(2\pi f t) + \eta_{it} + \epsilon_{it},$$
$$h_7(x_{it}, B) = (B_0 + B_1 \mathbf{x}_t)^2,$$

model 8:

$$y_{it} = \rho y_{i,t-1} + h_8(x_{it}, B) - \rho h_8(x_{i,t-1}, B) + (1-\rho)v_i + \beta_1 cos(2\pi f t) + \beta_2 sin(2\pi f t) + \eta_{it} + \epsilon_{it},$$
$$h_8(x_{it}, B) = B_0 + B_1^2 \mathbf{x}_t,$$

where t represents the month ($t = 1, 2, \cdots, 12$). The estimation of parameters of rival models are given in Appendix C. We consider 20 stations and estimate the parameters of rival models. The results for the value of the estimated parameters, the values of the Akaike information criterion, AIC, and mean-squared prediction error, $MSPE_h$ are reported in Table 18. Model 5 has the lowest value of the Akaike information criterion and $MSPE_h$, so we select model 5 as the optimal model based on the AIC and $MSPE_h$. The value of the Vuong's statistic for paired of rival models are computed and are given in Table 19. We compared the value of the Vuong's statistic with $z_{0.95}$. It shows that, the Vuong's test selects the model 5 against the models 6, 7 and 8. For more illustration see Figure 2.

Table 18: The value of estimated parameters of the proposed models, AIC and $MSPE_h$.

Area	$\hat{\mu}_\epsilon$	$\hat{\sigma}^2_\epsilon$	$\hat{\sigma}^2_v$	$\hat{\rho}$	$\hat{B}_0$	$\hat{B}_1$	$\hat{B}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$MSPE_h$	AIC
model 5											
A_1	1.9187	71.0731	1.0020	0.2908	342.1848	0.0594	3.4816	-75.9189	-14.7093	59.2898	132.0182
A_2	-10.4183	61.7377	1.0046	0.7012	461.0058	0.0226	-1.5092	-152.2698	22.4084	51.5056	128.9203
A_3	-15.2980	67.6542	1.0050	0.0326	423.0320	0.1284	0.1443	-148.9435	-107.3739	60.9125	130.9337
A_4	-24.0786	75.7197	1.0028	0.3484	407.4171	0.0337	1.2014	-102.0146	-25.0077	59.1198	133.4115
A_5	-17.1753	107.9227	1.0021	0.5145	394.2252	0.0375	3.5051	-101.7721	3.4047	88.8424	141.2078
A_6	-18.2315	68.4024	1.0041	0.5041	389.0683	0.0426	-1.3888	-137.0408	-64.0718	57.4172	131.1756
A_7	26.6798	75.7752	1.0006	0.7475	214.0461	0.7079	2.2683	-31.5275	15.3993	64.5003	133.4276
A_8	-16.9488	66.7008	1.0055	0.4792	406.0779	0.0638	-3.5918	-159.0095	-60.8232	61.2171	130.6214
A_9	-4.2077	29.5362	1.0005	0.2492	370.9456	0.0201	-1.0771	-35.3921	-33.0759	24.7770	112.7002
A_{10}	-17.2914	71.1059	1.0052	0.5263	411.8578	0.0257	-2.2868	-171.9353	-31.4372	62.0042	132.0284
A_{11}	-4.8302	67.3359	1.0048	0.0709	430.4085	0.1423	-0.3090	-125.4301	-105.9344	60.4263	130.8299
A_{12}	-2.4132	64.4585	1.0010	0.4725	441.0911	0.0171	-0.5926	-81.4573	4.9340	52.1409	129.8691
A_{13}	-17.8612	88.8500	1.0012	0.5673	377.3286	0.0341	0.7886	-80.8138	-42.2773	78.5286	136.9295
A_{14}	-4.6259	42.5290	1.0016	0.3513	353.8037	0.0595	-7.9469	-132.5470	-44.5539	37.6110	120.7207
A_{15}	-19.5667	96.5480	1.0023	0.3924	334.7104	0.0056	1.1666	-112.1326	21.2367	79.0026	138.7575
A_{16}	-0.8188	93.3341	1.0018	0.6027	104.7379	0.0366	13.8020	-199.5722	-12.9799	80.8764	138.0127
A_{17}	-12.6589	88.1615	1.0022	0.5867	427.2896	0.0740	4.9442	-49.9788	7.7095	74.0457	136.7584
A_{18}	-3.1888	40.4285	1.0003	0.5783	222.0354	0.0196	-1.8856	-70.5844	18.5880	32.2668	119.6064
A_{19}	17.4236	74.6146	1.0033	0.0062	467.4190	0.1259	1.0214	-105.5198	-109.4570	62.9980	133.0880
A_{20}	-24.2551	78.8927	1.0025	0.3253	421.8283	0.0350	0.6596	-90.1408	-28.2774	63.9077	134.3146
model 6											
A_1	66.8575	86.9660	1.0024	0.2754	3483.407	9.5451	3.9156	-66.4423	-12.4417	73.9964	136.4580
A_2	124.0467	142.0804	1.0067	0.6895	19214.245	41.6833	-0.3113	-153.0393	18.0855	154.2365	147.2573
A_3	173.4722	155.8472	1.0055	0.0606	2337.049	5.4486	1.1139	-147.8048	-98.7025	177.7214	149.2919
A_4	90.3335	119.0235	1.0034	0.3537	10133.650	25.1230	1.9094	-107.0606	-27.6311	98.2557	143.3617
A_5	77.1645	137.8827	1.0031	0.5217	10657.156	27.0985	4.8698	-102.4218	4.4747	127.2231	146.5975
A_6	158.0147	144.6638	1.0090	0.4627	12425.896	32.6025	1.7127	-142.0282	-76.6495	168.1643	147.6537
A_7	34.3345	96.7327	1.0011	0.9603	136.543	-90.4442	2.3790	-65.5251	65.3576	84.3604	138.7996
A_8	164.0492	151.2499	1.0080	0.4837	5155.441	12.6013	-1.8729	-157.2786	-59.0947	169.2648	148.6332
A_9	42.6216	45.4470	1.0009	0.1663	14551.489	38.3968	-0.3936	-28.6109	-34.4576	45.0934	122.1807
A_{10}	167.2739	161.0366	1.0074	0.5062	14591.831	35.4539	-1.1546	-173.3892	-38.0151	184.1331	150.0125
A_{11}	142.3582	132.6306	1.0048	0.1673	2184.523	4.9041	0.4774	-118.2649	-84.7623	148.8282	145.7431
A_{12}	73.0683	93.7851	1.0018	0.4726	26341.038	59.8496	0.3107	-82.9119	4.5485	87.1827	138.1188
A_{13}	81.9809	115.8527	1.0024	0.5786	11881.361	31.5720	2.1470	-82.1123	-39.9244	110.6551	142.7677
A_{14}	147.4049	119.8488	1.0020	0.3722	6755.972	19.5264	-5.2868	-133.7007	-42.5760	158.1186	143.5137
A_{15}	73.6716	129.7617	1.0014	0.6684	618.664	2.4526	1.3992	-59.4773	51.4741	128.1794	145.2620
A_{16}	193.2505	187.4752	1.0025	0.6085	4641.106	44.7780	14.7927	-199.3812	-11.8081	227.1216	153.3569
A_{17}	23.4576	94.9370	1.0019	0.6043	3414.505	7.6759	5.3180	-44.4767	11.3806	72.2080	138.3873
A_{18}	54.4218	73.0876	1.0005	0.5706	10737.720	48.3468	-1.3867	-70.8986	17.6450	75.1857	132.6331
A_{19}	146.8430	131.1303	1.0041	0.0417	2974.678	6.0751	-0.1529	-102.8173	-99.2635	151.9577	145.4929
A_{20}	71.4327	109.3710	1.0029	0.3569	9048.199	21.2534	1.3335	-88.6786	-23.9254	89.3141	141.5011
model 7											
A_1	13.3769	112.9682	3.6105	0.9432	21.8647	-0.2226	-	-68.9209	76.6311	94.1476	142.2130
A_2	122.8686	143.2448	1.3853	0.6907	21.4185	-0.1448	-	-153.5394	20.9175	157.0632	147.4369
A_3	118.4313	145.0992	1.0743	0.4552	20.7101	-0.5334	-	-132.9445	-47.3301	149.8676	147.7198
A_4	25.2087	107.8614	3.3883	0.7379	21.6391	0.2718	-	-69.3522	33.7799	85.0728	141.1953
A_5	24.0131	138.8384	1.8810	0.7781	21.8057	-0.3740	-	-83.0316	53.5111	111.4818	146.7495
A_6	176.8311	155.0814	1.1790	0.3419	19.4666	-0.2911	-	-140.4421	-109.7453	191.5638	149.1835
A_7	19.9269	122.1358	37.4739	0.9297	20.2987	0.0960	-	-68.0829	44.5703	100.2305	143.9296
A_8	198.8566	167.4801	1.1021	0.3590	19.7213	-0.3583	-	-170.7876	-91.8821	203.3231	150.8757
A_9	39.4605	47.9579	10.7092	0.5352	19.1477	-0.0705	-	-39.4871	-15.7095	46.0352	123.3637
A_{10}	157.7508	159.6241	1.4480	0.5352	20.3092	-0.1228	-	-170.9793	-26.7695	182.2478	149.8187
A_{11}	151.3467	152.8629	1.0907	0.3109	20.3598	-0.5215	-	-134.2323	-90.7629	176.5648	148.8666
A_{12}	75.8920	95.1471	4.6810	0.5101	20.5819	-0.1114	-	-81.4521	7.0988	90.7725	138.4360
A_{13}	67.1025	113.0988	2.0976	0.5962	19.9628	-0.3676	-	-71.5098	-34.1392	101.2232	142.2384
A_{14}	147.6842	117.7734	1.1233	0.4154	18.2618	-0.0617	-	-129.6435	-41.6546	161.0651	143.1294
A_{15}	12.7725	135.6658	82.7806	0.6222	20.0763	-0.0908	-	-81.1222	55.8668	96.9549	146.2409
A_{16}	67.1988	187.6723	1.0361	0.9490	17.4525	-0.5949	-	-162.9472	66.5432	153.0725	153.3800
A_{17}	27.1142	132.1854	4.2708	0.9475	24.3445	0.2740	-	-90.5663	97.6048	110.3625	145.6692
A_{18}	53.6700	73.5315	1.3317	0.5737	14.8700	-0.0382	-	-71.2646	18.7869	76.5379	132.7664
A_{19}	203.0199	162.4459	1.0739	0.1050	20.6179	-0.4821	-	-128.2581	-128.7800	203.9011	150.2042
A_{20}	38.6512	109.3047	3.9149	0.5572	21.2221	0.2554	-	-72.0279	6.6018	91.9144	141.4877
model 8											
A_1	2138.357	875.371	1.0025	-2.2530	8.6982	0.2886	-	-1074.723	-1293.469	2138.357	187.258
A_2	8384.010	163.327	1.0280	52.7765	116.2275	0.0000	-	-9298.953	-3485.304	8384.010	808.862
A_3	-1572.336	462.843	1.0315	-0.5223	94.7546	-0.0551	-	1587.332	968.520	1572.336	173.239
A_4	2242.644	674.417	1.0296	-1.2389	63.3879	0.0866	-	500.689	-618.948	608.412	181.521
A_5	1434.597	266.305	1.0247	0.2297	38.6669	0.8517	-	-1372.931	-192.425	1434.597	161.078
A_6	4219.516	198.948	1.0264	0.4585	89.6206	0.7102	-	-690.885	584.773	421.951	154.663
A_7	3522.695	397.982	1.0258	-0.2801	-44.7429	0.9713	-	-3134.716	-1427.093	3522.695	169.917
A_8	1852.164	547.920	1.0279	17.5921	99.8305	0.0000	-	-2239.246	-6074.974	1852.164	784.834
A_9	-4511.269	785.987	1.0021	1.2924	86.3180	-0.5539	-	949.711	891.810	451.126	134.232
A_{10}	4312.271	104.033	1.0251	35.1843	105.2731	0.0000	-	-5000.059	-1694.167	4312.271	798.939
A_{11}	2116.878	726.433	1.0291	3.0396	83.1374	-0.2491	-	-1886.856	-1781.717	2116.878	183.155
A_{12}	1912.218	407.886	1.0107	21.3001	94.0504	0.0000	-	-2361.568	-9186.575	1912.218	525.056
A_{13}	4540.266	153.329	1.0080	0.8812	71.9527	0.2867	-	-570.491	120.864	454.026	148.933
A_{14}	1000.818	151.239	1.0076	1.6136	89.9181	-0.3206	-	-773.515	-1026.617	1000.818	148.631
A_{15}	4834.978	217.428	1.0018	-7.0543	20.4921	0.0932	-	-1974.018	-3616.533	4834.978	207.274
A_{16}	1123.206	242.333	1.0035	0.9574	18.9609	0.4748	-	-264.452	188.450	206.038	159.003
A_{17}	1506.856	802.716	1.0058	-1.1541	-7.9529	0.4906	-	-908.421	-889.276	1506.856	185.352
A_{18}	-2015.494	803.646	1.0021	0.4799	64.7153	-0.0310	-	205.721	206.209	201.549	134.721
A_{19}	2190.035	520.378	1.0251	17.59219	105.7737	0.0000	-	-2553.320	-8004.597	2190.035	783.699
A_{20}	-2727.249	143.929	1.0278	-3.7867	79.6964	-0.0026	-	3373.206	1037.288	2727.249	198.198

Table 19: The values of Vuong's test.

Area	$\begin{cases} H_f : model5 \\ H_g : model6 \end{cases}$	$\begin{cases} H_f : model5 \\ H_g : model7 \end{cases}$	$\begin{cases} H_f : model6 \\ H_g : model7 \end{cases}$	$\begin{cases} H_f : model5 \\ H_g : model8 \end{cases}$	$\begin{cases} H_f : model6 \\ H_g : model8 \end{cases}$	$\begin{cases} H_f : model7 \\ H_g : model8 \end{cases}$
A_1	1.6390	2.9782	1.4693	12.3411	10.5543	10.7780
A_2	5.1341	5.0265	0.2116	129.2823	118.9785	113.0560
A_3	5.3073	4.3315	-0.7422	9.8919	10.6242	7.7196
A_4	2.9020	1.8894	-0.6150	10.3615	10.3019	8.2226
A_5	2.7919	3.0962	0.0761	4.4456	2.8559	2.7737
A_6	4.4773	4.5642	1.5706	4.8870	2.4643	1.5902
A_7	1.0955	4.6623	0.9838	12.7219	6.7424	8.0283
A_8	6.3809	6.7705	1.7242	183.0188	173.3721	171.6382
A_9	3.0573	3.6653	0.4258	6.9338	2.9040	2.7786
A_{10}	5.5487	5.3846	-0.2073	167.8561	151.5434	135.0003
A_{11}	3.9492	4.2099	2.0389	12.0984	9.5226	8.7918
A_{12}	2.6801	2.6834	0.5264	78.8859	74.1305	75.6960
A_{13}	2.1520	2.2558	-0.9792	3.8591	1.5696	1.7473
A_{14}	5.6003	5.0755	-0.4295	7.0776	1.0294	1.0492
A_{15}	2.1408	2.8877	0.3835	14.7927	11.5978	12.1650
A_{16}	5.3420	6.0740	0.0108	6.1486	1.4515	1.5128
A_{17}	0.7844	3.9279	2.9402	9.8783	8.4765	7.7231
A_{18}	3.0813	2.9065	0.1420	3.1661	1.0393	1.4961
A_{19}	2.7839	3.7641	2.9553	130.8784	169.3102	178.6782
A_{20}	2.0831	1.7959	-0.0070	12.1922	12.4220	10.8482

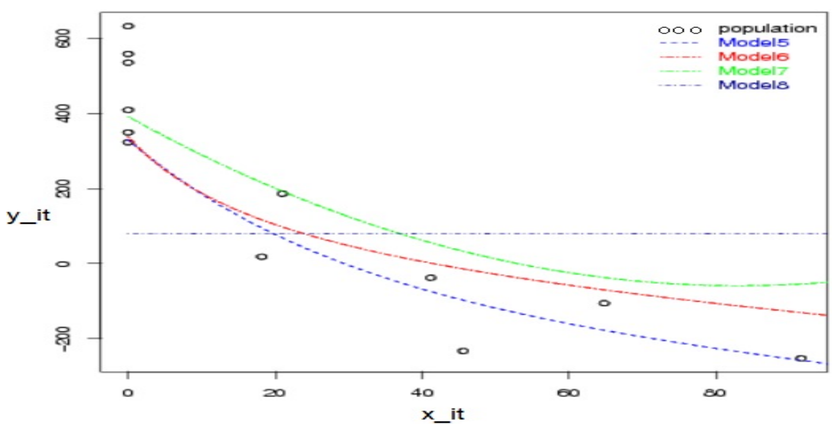

Figure 2: The curve of competing models for station 20.

The X-axis represents the monthly sum of snowfall and the Y-axis is the value of the monthly sum of mean temperatures. It shows that the estimated model 5 is better than estimated model 6 and 7. The estimated model 8 creates the worst fit.

The values of forecasts for 20 stations are given in Table 20. The results show that, the model 5 has the closest prediction to the true values.

Since Vuong's test selects estimated model 5 as optimal model and estimated model 5 has the lowest value $MSPE_h$ and AIC, so we select model 5 as an optimal model.

Table 20: The values of prediction based on the rival models for 20 stations.

	True	model 5	model 6	model 7	model 8
A_1	452.5	455.8231	333.3080	460.0273	-2634.217
A_2	736.3	743.5005	496.6321	497.2091	-6758.108
A_3	607.5	611.7354	262.4199	371.5804	1.685.666
A_4	560.6	554.0434	344.7395	477.1373	-401.4504
A_5	719.1	541.9770	369.4762	459.8345	-1137.509
A_6	608.6	647.4270	322.2742	278.9192	18.8796
A_7	437.0	323.6091	297.1312	303.1851	-357.9229
A_8	648.4	682.2564	350.9758	282.2109	-1307.493
A_9	448.1	435.2690	347.0663	363.6853	919.6064
A_{10}	682.4	703.4517	364.6726	379.7896	-3249.767
A_{11}	628.3	603.3244	324.8577	307.5046	-900.5252
A_{12}	660.9	574.6587	436.5525	436.2365	-1470.066
A_{13}	511.5	527.0924	345.6385	368.5153	-1.494507
A_{14}	567.4	560.1092	285.1896	290.7058	-349.4189
A_{15}	628.3	505.2589	424.7602	451.4870	-7224.866
A_{16}	577.0	504.5205	148.3215	354.1947	316.4679
A_{17}	602.6	530.7318	467.3436	509.7241	-1867.557
A_{18}	300.2	323.3473	218.3106	218.4046	462.5413
A_{19}	663.6	614.6202	357.3427	272.4461	-1667.324
A_{20}	558.0	550.8766	376.0042	438.5052	1795.151

4.1 Manitoba Data

We use a monthly dataset of climate change Canadian province of Manitoba from January 2022 to December 2022. The province consisted 79 stations. We selected 9 stations from Manitoba for analysis.

Table 21: Nine selected weather stations from Manitoba used for model evaluation in 2022.

Parameter	Station	Province
A_1	BALDUR	MANITOBA
A_2	BRANDON A	MANITOBA
A_3	ELKHORN 2 EAST	MANITOBA
A_4	INDIAN BAY	MANITOBA
A_5	MARQUETTE	MANITOBA
A_6	GILLAM A	MANITOBA
A_7	STONY MOUNTAIN	MANITOBA
A_8	THOMPSON A	MANITOBA
A_9	VIRDEN WATER	MANITOBA

The stations, RANDON A and ELKHORN 2 EAST, don't follow of rival models. So, we considered the 7 stations as dataset and model 5 to model 8 as the rival models. We estimate the parameters of rival models and reported the value of the estimated parameters, the values of the Akaike Information Criterion, AIC, and Mean-Squared Prediction Error, $MSPE_h$ in Table 22. The results shows that the model 5 has the lowest value of the AIC and $MSPE_h$.

The values of Vuong's statistic for paired of rival models are computed and are given in Table 23. We compared the value of the Vuong's statistic with $z_{0.95}$. It shows that, the Vuong's test selects the model 5 is better than the model 6, model 7 and model 8. For more illustration see Figure 3. The X-axis represents the monthly sum of snowfall and the Y-axis is the value of the monthly sum of mean temperatures. It shows that the estimated model 7 and 8 do not create proper fits. The estimated model 5 is better than estimated model 6.

Table 22: The value of estimated parameters of the proposed models, AIC and $MSPE_h$.

Area	$\hat{\mu}_\epsilon$	$\hat{\sigma}^2_\epsilon$	$\hat{\sigma}^2_v$	$\hat{\rho}$	$\hat{B}_0$	$\hat{B}_1$	$\hat{B}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$MSPE_h$	AIC
					model 5						
A_1	31.6943	154.4794	0.9999	0.6986	235.8562	0.3713	6.5300	-127.2852	51.5260	132.5765	168.5330
A_2	4.3929	152.0204	0.9999	0.6628	291.8761	0.5314	4.4717	-148.4742	17.5341	116.7794	168.5734
A_3	-39.6928	145.7110	0.9999	0.7189	484.2704	0.2774	-4.6733	-119.0389	42.2992	121.3695	166.5067
A_4	-45.0781	150.2976	0.9999	0.8053	228.8458	0.9758	2.1657	-178.8337	55.9895	112.5609	168.2882
A_5	-37.3118	156.9454	0.9999	0.7482	451.5861	0.6247	-6.2498	-152.0756	61.5824	120.8987	169.6704
A_6	-12.7622	158.2032	0.9999	0.5180	261.4957	0.1860	6.5292	-132.4267	-2.6369	125.3152	170.4368
A_7	29.8748	139.0192	0.9999	0.7013	271.6336	0.5225	5.9623	-123.4351	43.4789	115.5249	169.0752
					model 6						
A_1	19.5450	168.5600	1.0006	1.0222	-38.5245	-6.8775	9.3756	-86.1456	148.6225	145.1353	293.1353
A_2	55.5381	145.6261	1.0004	1.0154	119.8440	-7.3203	15.0007	-112.1533	107.0340	125.4583	291.1970
A_3	187.6937	214.9857	1.0008	0.7706	-46.1149	-2.0188	4.2474	-192.9496	61.3186	227.1888	278.8069
A_4	106.1035	213.2422	1.0008	0.7526	-79.6246	-0.3576	1.5932	-183.8670	28.6667	187.8144	281.7978
A_5	87.4450	187.2937	1.0007	0.8855	-313.3551	-4.3044	12.6649	-136.8757	110.4395	158.7462	293.7598
A_6	197.5773	208.2014	1.0073	0.6230	-386.4628	-11.4070	7.3587	-179.6126	8.4709	257.5382	283.9281
A_7	37.7289	175.7815	1.0007	1.0182	-104.9378	-6.5483	9.0782	-101.4009	147.0155	133.0896	293.8134
					model 7						
A_1	54.1816	194.9948	1.0893	0.7912	21.8083	-0.4027	-	-129.6200	72.0297	156.5259	283.9596
A_2	127.4626	206.6277	1.0633	0.6788	19.5903	-0.4814	-	-180.6527	-6.9147	207.3007	276.3817
A_3	35.1331	168.8286	1.0338	0.7723	22.6346	-1.3232	-	-101.5328	56.3047	126.4901	282.9573
A_4	80.6278	194.2301	1.0318	0.9937	-18.9443	0.6533	-	-158.2914	86.3684	144.3064	288.8826
A_5	65.3237	199.5453	1.0135	0.8659	22.2364	-2.7610	-	-157.9121	61.8637	158.9496	286.2566
A_6	83.1379	232.8209	1.3157	0.9406	18.3420	-0.2466	-	-179.4252	89.4267	180.5132	289.1396
A_7	62.4507	178.9990	1.0892	0.8040	22.2793	-0.4143	-	-129.2819	69.2302	146.8072	285.8954
					model 8						
A_1	1285.8728	600.2341	1.0008	-0.7133	16.2113	0.5525	-	-812.1815	-638.5573	1285.8728	272.9445
A_2	157.9988	270.5531	1.0008	1.0966	52.4400	0.2931	-	-367.8513	293.1938	228.5064	294.1917
A_3	-255.8197	238.9348	1.0108	0.9785	55.9162	-1.1720	-	286.6726	23.7547	293.9770	286.0954
A_4	-942.0899	1459.5359	1.0010	4.6918	-35.9030	-0.1803	-	-144.8467	2024.7534	1472.1466	356.7325
A_5	-279.4147	277.1970	1.0014	1.0922	43.0830	-2.0707	-	245.6905	90.8301	329.0517	295.9128
A_6	-952.0567	1492.9107	1.0019	4.6436	-35.8374	-0.1419	-	-319.5895	2204.9663	1500.8729	357.9214
A_7	1236.4294	549.3784	1.0009	-0.5195	24.3516	0.5775	-	-806.1113	-581.1634	1236.4294	260.1735

The values of forecasts for 7 stations are given in Table 24. The results show that, the model 5 has the closest prediction to the true values, except for two cases.

Since Vuong's test selects model 5 as an optimal model and model 5 has the lowest of $MSPE_h$ and AIC, so we select model 5 as an optimal model.

Table 23: The values of Vuong's test.

Area	$\begin{cases} H_f : model5 \\ H_g : model6 \end{cases}$	$\begin{cases} H_f : model5 \\ H_g : model7 \end{cases}$	$\begin{cases} H_f : model6 \\ H_g : model7 \end{cases}$	$\begin{cases} H_f : model5 \\ H_g : model8 \end{cases}$	$\begin{cases} H_f : model6 \\ H_g : model8 \end{cases}$	$\begin{cases} H_f : model7 \\ H_g : model8 \end{cases}$
A_1	64.3599	36.4858	42.2874	6.9622	6.5305	27.86712
A_2	79.9658	33.4311	35.8403	3.2575	2.4574	24.53049
A_3	57.8962	32.3729	29.5292	3.0853	0.8559	25.17017
A_4	52.3257	32.6215	22.8248	8.9403	6.6134	19.15768
A_5	64.3224	29.1118	24.0156	3.7295	1.6479	25.43459
A_6	48.3754	31.4229	27.7917	14.3239	9.0256	20.17827
A_7	69.5053	27.8776	28.5469	5.4550	4.5172	28.90255

Table 24: The values of prediction based on the rival models for 7 stations.

Area	True	model 5	model 6	model 7	model 8
A_1	568.7	412.4899	312.88846	265.5610	-1213.5078
A_2	557.6	454.9482	275.38379	176.4350	167.1877
A_3	608.5	508.7173	140.60117	329.2660	206.4697
A_4	224.8	493.6710	157.61064	229.9578	1403.1424
A_5	616.7	516.9600	242.27486	251.1702	620.8599
A_6	398.6	491.1830	62.92666	120.8465	1395.2991
A_7	577.0	434.5993	319.37600	283.6450	-1122.0826

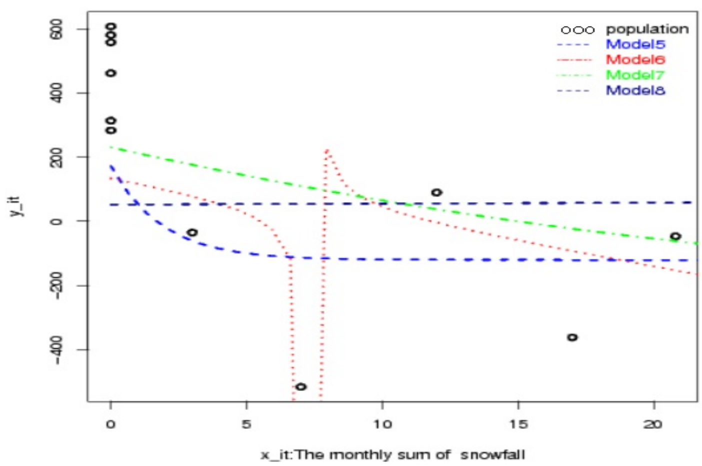

Figure 3: The curve of competing models for station 7 of Manitoba data.

5 Conclusion

In this paper, we make a number of contributions to the literature that relates to an extension of the basic Fay and Herriot model involving non-linear regression model, autocorrelated random effects and sampling error for small area estimation with utilizing both time series and cross -sectional data. First, we estimated the parameters of model and variance of random effects based on the EM algorithm. This is important, because the standard methods of variance component estimation used in the Fay and Herriot model for small areas produce a negative or zero estimate. Second, we proposed a test for equivalence of the model in all small areas. Also, we examined the properties of estimators and model selection statistic using simulations which validate our theoretical results. In summary, the theoretical derivations and the simulations support the use of EM algorithm and the use of proposed model selection test as a useful tool in a practitioner's toolbox for empirical analysis of small area. The simulation results show that the performance of the semi-parametric model is better than other rival models. The Vuong's test selected the semi-parametric model and the true model (model 1) as equivalent models. It also results that the semi-parametric model is better than other rival models.

Conflict of interest: The Authors declare that they have no conflict of interest.

This research was completed at University of Manitoba, Canada, where the first author was in sabbatical leave by the K.N. Tossi University of Technology, Tehran, Iran.

References

Fay, R. E., and Herriot, R. A. (1979). Estimates of income for small places: an application of James–Stein procedures to census data. *Journal of the American Statistical Association*, **74**, 269–277.

Jiang, J., Rao, J. S., Gu, Z., and Nguyen, T. (2008). Fence methods for mixed model selection. *The Annals of Statistics*, **36**, 1669–1692.

Jiang, J., Nguyen, T., and Rao, J. S. (2010). Fence method for non-parametric small area estimation. *Survey Methodology*, **36**, 3–11.

Jiang, J., Nguyen, T., and Rao, J. S. (2011). Best predictive small area estimation. *Journal of the American Statistical Association*, **106**(494), 732–745.

Jiang, J., Nguyen, T., and Lahiri, P. (2018). A unified Monte-Carlo jackknife for small area estimation after model selection. *Annals of Mathematical Sciences and Applications*, **3**, 405–438.

Lahiri, P., and Rao, J. N. K. (1995). Robust estimation of mean squared error of small area estimators. *Journal of the American Statistical Association*, **82**, 758–766.

Nielsen, S. F. (2000). The stochastic EM algorithm: Estimation and asymptotic results. *Bernoulli*, **6**(3), 457–489.

Prasad, N. G. N., and Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, **85**, 163–171.

Rao, J. N. K., and Molina, I. (2015). Empirical Bayes and hierarchical Bayes estimation of poverty measures for small areas. In M. Pratesi (Ed.), *Analysis of Poverty Data by Small Area Methods*. John Wiley & Sons, Inc.

Rao, J. N. K., and Molina, I. (2015). *Small Area Estimation* (2nd ed.). Wiley, New York. https://doi.org/10.1002/9781118735855.

Rao, J. N. K., and Yu, M. (1992). Small area estimation by combining time series and cross-sectional data. *Proceedings of the Section on Survey Research Method, American Statistical Association*, 1–9.

Rao, J. N. K., and Yu, M. (1994). Small area estimation by combining time series and cross-sectional data. *Canadian Journal of Statistics*, **22**, 511–528.

Sayyareh, A. (2012). Inference after separated hypotheses testing: an empirical investigation for linear models. *Journal of Statistical Computation and Simulation*, **82**(9), 1275–1286.

Schenker, N., and Welsh, A. H. (1987). Asymptotic results for multiple imputation. *Annals of Statistics*, **16**, 1550–1566.

Shokoohi, F., and Torabi, M. (2018). Semi-parametric small-area estimation by combining time-series and cross-sectional data methods. *Australian & New Zealand Journal of Statistics*, **60**(3), 323–342. https://doi.org/10.1111/anzs.12234.

Sugasawa, S., Kawakubo, Y., and Datta, G. S. (2019). Observed best selective prediction in small area estimation. *Journal of Multivariate Analysis*, **173**, 383–392.

Sugasawa, S., and Kubokawa, T. (2020). Small area estimation with mixed models: a review. *Japanese Journal of Statistics and Data Science*, **3**, 693–720. https://doi.org/10.1007/s42081-020-00076-x.

Sundberg, R. (2019). *Statistical Modelling by Exponential Families*. Cambridge University Press.

Torabi, M., and Shokoohi, F. (2014). Hierarchical Bayes estimation in small area estimation using cross-sectional and time-series data. *Journal of Statistical Computation and Simulation*, **84**(3), 605–613. https://doi.org/10.1080/00949655.2012.721365.

Torabi, M., and Shokoohi, F. (2015). Non-parametric generalized linear mixed models in small area estimation. *Canadian Journal of Statistics*, **43**(1), 82–96. https://doi.org/10.1002/cjs.11236.

Vaida, F., and Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika*, **92**, 351–370.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, **57**(2), 307–333.

Wolter, K. (1985). *Introduction to Variance Estimation*. Springer-Verlag, New York.

Appendix A

Each model satisfies the Vuong's conditions, Vuong (1989).

ASSUMPTION A1: For every ϑ in V (parameter space), the distribution $F(y;\vartheta)$ has a Radon-Nikodym density $f(y;\vartheta)$ which is strictly positive.

ASSUMPTION A2: V is a compact subset of $\mathfrak{R}^q$ and the density $f(y;\vartheta)$ is continuous in ϑ.

Given Assumption A1 and A2, for all T there exists a measurable MLE $\hat{\vartheta}$. It ensures that a MLE always exists, but does not say anything about its uniqueness.

ASSUMPTION A3: (a) For all y, $|\log f(y;\vartheta)|$ is dominated by an integrable functions independent on ϑ. (b) The function $\int \log f(y;\vartheta)dF(y)$ has a unique maximum on V at ϑ_*.

The value ϑ_* is called the pesudo-true value of ϑ for the conditional distribution $F(y;\vartheta)$.

Assumption A3 (b) is the fundamental identification condition, Vuong (1989). When Assumption A3 (b) holds, we say that ϑ_* is globally identifiable.

ASSUMPTION A4: (a) For all y, $\log f(y;\vartheta)$ is twice continuously differentiable on V. (b) For all y, $\mid \frac{\partial \log f(y;\vartheta)}{\partial\vartheta} . \frac{\partial \log f(y;\vartheta)}{\partial\vartheta} \mid$ and $\mid \frac{\partial^2 \log f(y;\vartheta)}{\partial\vartheta\partial\vartheta'} \mid$ are dominated by integrable functions independent on ϑ.

Assumption A4 (b) ensures that the derivatives are appropriately dominated by integrable functions with respect to F, which ensures that $C_f(\vartheta) = E\left(\frac{\partial^2 \log f(Y;\vartheta)}{\partial\vartheta\partial\vartheta'}\right)$ and $B_f(\vartheta) = E\left(\frac{\partial \log f(Y;\vartheta)}{\partial\vartheta} . \frac{\partial \log f(Y;\vartheta)}{\partial\vartheta'}\right)$ are continuous in ϑ.

ASSUMPTION A5: (a) ϑ_* is an interior point of V. (b) ϑ_* is a regular point of $C_f(\vartheta)$.

In Assumption A5 (b), a regular point of matrix $C_f(\vartheta)$ is defined as a value for ϑ such that $C_f(\vartheta)$ has constant rank in some open neighborhood of ϑ.

ASSUMPTION A6: $\mid \frac{\partial \log f(y;\vartheta)}{\partial\vartheta} \mid$ and $\mid \frac{\partial^2 \log f(y;\vartheta)}{\partial\vartheta\partial\vartheta'} \mid$ are differentiable on V and these derivatives are appropriately dominated by functions integrable with respect to $F(y;\vartheta)$.

Appendix B

Proof of Theorem 2.2. From the Taylor expansion of $L_c(\vartheta^*_{it}; y_{i,t})$ around $\hat{\vartheta}_{it}$ we obtain

$$\begin{aligned}
L_c(\vartheta^*_{it}; y_{i,t}) &= L_c(\hat{\vartheta}_{it}; y_{i,t}) + \frac{\partial L_c(\vartheta_{it}; y_{i,t})}{\partial\vartheta_{it}}\Big|_{\vartheta_{it}=\hat{\vartheta}_{it}} \left(\vartheta_{it} - \hat{\vartheta}_{it}\right) \\
&+ \left(\frac{\partial^2 L_c(\vartheta_{it}; y_{i,t})}{\partial\vartheta_{it}\partial\vartheta'_{it}}\Big|_{\vartheta_{it}=\hat{\vartheta}_{it}}\right) \frac{(\vartheta_{it} - \hat{\vartheta}_{it})(\vartheta - \hat{\vartheta}_{it})'}{2} + O_p(1) \\
&= L_c(\hat{\vartheta}_{it}; y_{i,t}) + \left(\frac{\partial^2 L_c(\vartheta_{it}; y_{i,t})}{\partial\vartheta_{it}\partial\vartheta'_{it}}\Big|_{\vartheta_{it}=\hat{\vartheta}_{it}}\right) \frac{(\vartheta_{it} - \hat{\vartheta}_{it})(\vartheta_{it} - \hat{\vartheta}_{it})'}{2} + O_p(1).
\end{aligned}$$

Note that

$$\frac{1}{n}\left(\frac{\partial^2 L_c(\vartheta_{it}; y_{i,t})}{\partial\vartheta_{it}\partial\vartheta'_{it}}\Big|_{\vartheta_{it}=\hat{\vartheta}_{it}}\right) \xrightarrow{P} E_h\left\{\frac{\partial^2 \log f(y_{it}; \vartheta_{it})}{\partial\vartheta_{it}\partial\vartheta'_{it}}\right\} = A_{f^{\vartheta_{it}}},$$

where $A_{f^{\vartheta_{it}}}$ is the complete data information. So, we can write

$$L_c(\hat{\vartheta}_{it}; y_{i,t}) + \frac{n}{2}(\vartheta_{it} - \hat{\vartheta}_{it}) A_{f^{\vartheta_{it}}} (\vartheta_{it} - \hat{\vartheta}_{it})' + O_p(1).$$

Similarly, we have

$$
\begin{aligned}
L_c(\nu^*_{it}; y_{i,t}) &= L_c(\hat{\nu}_{it}; y_{i,t}) + \frac{\partial L_c(\nu_{it}; y_{i,t})}{\partial \nu_{it}}|_{\nu_{it}=\hat{\nu}_{it}} (\nu_{it} - \hat{\nu}_{it}) \\
&+ \left(\frac{\partial^2 L_c(\nu_{it}; y_{i,t})}{\partial \nu_{it} \partial \nu^{'}_{it}}|_{\nu_{it}=\hat{\nu}_{it}} \right) \frac{(\nu_{it} - \hat{\nu}_{it})(\nu - \hat{\nu}_{it})^{'}}{2} + O_p(1) \\
&= L_c(\hat{\nu}_{it}; y_{i,t}) + \left(\frac{\partial^2 L_c(\nu_{it}; y_{i,t})}{\partial \nu_{it} \partial \nu^{'}_{it}}|_{\nu_{it}=\hat{\nu}_{it}} \right) \frac{(\nu_{it} - \hat{\nu}_{it})(\nu_{it} - \hat{\nu}_{it})^{'}}{2} + O_p(1).
\end{aligned}
$$

and

$$
\frac{1}{n} \left(\frac{\partial^2 L_c(\nu_{it}; y_{i,t})}{\partial \nu_{it} \partial \nu^{'}_{it}}|_{\nu_{it}=\hat{\nu}_{it}} \right) \xrightarrow{P} E_h \left\{ \frac{\partial^2 \log f(y_{it}; \nu_{it})}{\partial \nu_{it} \partial \nu^{'}_{it}} \right\} = A_{f^{\nu_{it}}}.
$$

So, we can write

$$
L_c(\hat{\nu}_{it}; y_{i,t}) + \frac{n}{2}(\nu_{it} - \hat{\nu}_{it}) A_{f^{\nu_{it}}} (\nu_{it} - \hat{\nu}_{it})^{'} + O_p(1).
$$

Since $LR_c(\hat{\vartheta}_{it}, \hat{\nu}_{it}) = L_c(\hat{\vartheta}_{it}; y_{i,t}) - L^f_c(\hat{\nu}_{it}; y_{i,t})$, so we obtain

$$
LR_c(\hat{\vartheta}_{it}, \hat{\nu}_{it}) = LR_c(\vartheta^*_{it}, \nu^*_{it}) - \frac{n}{2}(\vartheta_{it} - \hat{\vartheta}_{it})^{'} A_{f^{\vartheta_{it}}} (\vartheta_{it} - \hat{\vartheta}_{it}) + \frac{n}{2}(\nu_{it} - \hat{\nu}_{it})^{'} A_{f^{\nu_{it}}} (\nu_{it} - \hat{\nu}_{it}) + O_p(1).
$$

To prove (i), we note that $n^{-\frac{1}{2}} LR_c(\hat{\vartheta}_{it}, \hat{\nu}_{it}) = n^{-\frac{1}{2}} LR_c(\vartheta^*_{it}, \nu^*_{it}) + o_p(1)$. So we have

$$
n^{-\frac{1}{2}} LR_c(\hat{\vartheta}_{it}, \hat{\nu}_{it}) - n^{\frac{1}{2}} E_h \left\{ \log \left(\frac{f(Y; \vartheta^*_{it})}{f(Y; \vartheta^*_{it})} \right) \right\} = n^{\frac{1}{2}} \left[\frac{1}{n} LR_c(\vartheta^*_{it}, \nu^*_{it}) - E_h \left\{ \log \left(\frac{f(Y; \vartheta^*_{it})}{f(Y; \vartheta^*_{it})} \right) \right\} \right] + o_p(1).
$$

Using multivariate central limit Theorem

$$
n^{-\frac{1}{2}} LR_c(\hat{\vartheta}_{it}, \hat{\nu}_{it}) - n^{\frac{1}{2}} E_h \left\{ \log \left(\frac{f(Y; \vartheta^*_{it})}{f(Y; \vartheta^*_{it})} \right) \right\} \xrightarrow{D} N(0, \omega^2_*).
$$

To prove (ii), if $f(y_{i,t}; \vartheta^*_{it}) = f(y_{i,t}; \nu^*_{it})$ then $LR_c(\vartheta^*_{it}, \nu^*_{it}) = 0$ and

$$
\begin{aligned}
LR_c(\hat{\vartheta}_{it}, \hat{\nu}_{it}) &= -\frac{n}{2}(\vartheta_{it} - \hat{\vartheta}_{it})^{'} A_{f^{\vartheta_{it}}} (\vartheta_{it} - \hat{\vartheta}_{it}) + \frac{n}{2}(\nu_{it} - \hat{\nu}_{it})^{'} A_{f^{\nu_{it}}} (\nu_{it} - \hat{\nu}_{it}) + o_p(1) \\
&= \begin{pmatrix} \vartheta_{it} - \hat{\vartheta}_{it} & \nu_{it} - \hat{\nu}_{it} \end{pmatrix} Q \begin{pmatrix} \vartheta_{it} - \hat{\vartheta}_{it} \\ \nu_{it} - \hat{\nu}_{it} \end{pmatrix} + o_p(1) \xrightarrow{D} M_{p+q}(., \lambda^*),
\end{aligned}
$$

where $Q = \begin{pmatrix} -A_{f^{\vartheta_{it}}} & 0 \\ 0 & A_{f^{\nu_{it}}} \end{pmatrix}$, λ^* are the eigenvalues of $Q\Sigma_{V^*}$ and

$$
\Sigma_{V^*} = I(V^*)^{-1} \left[2I - (I + F(V^*))^{-1} \right].
$$

□

Proof of Corollary 1. Straightforward from Theorem 2.2. □

Appendix C

Proof. From model (4.1), we have that $\mathbf{y}_c = \begin{pmatrix} \mathbf{y}_t \\ \mathbf{v} \end{pmatrix} \sim N_{2M}\left(\begin{bmatrix} \boldsymbol{\mu}_y \\ 0 \end{bmatrix}, \begin{bmatrix} V_y & \sigma_v^2 I_M \\ \sigma_v^2 I_M & \sigma_v^2 I_M \end{bmatrix}\right)$, where $\mathbf{y}_t$ is observed data, $\mathbf{v}$ is vector of unobserved or missing data, $\boldsymbol{\mu}_y = E(\mathbf{Y}_t)$, $\mathbf{V}_y = Var(\mathbf{Y}_t) = \mathbf{G} + \Sigma + \left(1 - \rho\rho'\right)^{-1} S$ and

$$E(Y_{it}) = (1-\rho)^{-1}\left(h_i(x_{it}, B) - \rho h_i(x_{i,t-1}, B) + \beta_1 cos(2\pi ft) + \beta_2 sin(2\pi ft)\right),$$

and $Var(Y_{it}) = \sigma_v^2 + \sigma_{e_i}^2 + (1-\rho^2)\sigma_{\epsilon_i}^2$. The complete log- likelihood function can be expressed as

$$\begin{aligned} l(B, \sigma_v^2, \rho; y_c) &= l(B, \sigma_v^2, \rho; y|v) + l(\sigma_v^2, \rho; v) \\ &= C - \frac{1}{2}\log|\varpi| - \frac{1}{2}\mathbf{z}_t'\varpi^{-1}\mathbf{z}_t - \frac{1}{2}\log|\sigma_v^2 I_M| - \frac{1}{2\sigma_v^2}\mathbf{v}'\mathbf{v}, \end{aligned}$$

where C is a constant that is independent of σ_v^2, $\varpi = S + \mathcal{S}$ and

$$\mathbf{z}_t = \mathbf{y}_t - \boldsymbol{\rho}\mathbf{y}_{t-1} - \mathbf{h}(\mathbf{x}_t, \mathbf{B}) + \boldsymbol{\rho}\mathbf{h}(\mathbf{x}_{t-1}, \mathbf{B}) - (1-\boldsymbol{\rho})I_M\mathbf{v} - \beta_1 cos(2\pi ft) - \beta_2 sin(2\pi ft).$$

Let $\theta^{(0)} = \left(\mathbf{B}^{(0)}, \sigma_v^{2(0)}, \boldsymbol{\rho}^{(0)}, \beta_1^{(0)}, \beta_2^{(0)}\right)$ be starting values of $\theta = \left(\mathbf{B}, \sigma_v^2, \boldsymbol{\rho}, \beta_1, \beta_2\right)$, so

$$\begin{aligned} Q_1 &= E\left(l(B, \sigma_v^2, \rho; y_c)\right) \\ &= C - \frac{1}{2}\log|\varpi| - \frac{1}{2}E\left(\mathbf{z}_t'\boldsymbol{\varpi}^{-1}\mathbf{z}_t \mid \mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right) - \frac{1}{2}\log|\sigma_v^2 I_M| - \frac{1}{2\sigma_v^2}E\left(\mathbf{v}'\mathbf{v}|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right). \end{aligned} \quad (5.1)$$

The third term of Equation (5.1) is equal to

$$E\left(\mathbf{z}_t'\boldsymbol{\varpi}^{-1}\mathbf{z}_t|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right) = E\left(\mathbf{z}_t'|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right)\boldsymbol{\varpi}^{-1}E\left(\mathbf{z}_t|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right) + tr\left[\boldsymbol{\varpi}^{-1}Var\left(\mathbf{z}_t|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right)\right], \quad (5.2)$$

where

$$\tilde{\mathbf{z}}_1 = \mathbf{y}_t - \boldsymbol{\rho}^{(0)}\mathbf{y}_{t-1} - \mathbf{h}(\mathbf{x}_t, \mathbf{B}^{(0)}) + \boldsymbol{\rho}^{(0)}\mathbf{h}(\mathbf{x}_{t-1}, \mathbf{B}^{(0)}) - (1-\boldsymbol{\rho}^{(0)})\tilde{b}_1 - \beta_1 cos(2\pi ft) - \beta_2 sin(2\pi ft), \quad (5.3)$$

$$\tilde{b}_1 = E\left(\mathbf{v}|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right) = \sigma_v^{2(0)}V_y^{-1}(\mathbf{y}_t - \boldsymbol{\mu}_y^{(0)}),$$

and

$$Var\left(\mathbf{z}_t|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right) = \sigma_v^{2(0)}(1-\boldsymbol{\rho}^{(0)})^2(I_M - \boldsymbol{\sigma}_v^{2(0)}\mathbf{V}_y^{-1}). \quad (5.4)$$

By substituting Equations (5.3) and (5.4) in Equation (5.2), we have

$$E\left(\mathbf{z}_t'\boldsymbol{\varpi}^{-1}\mathbf{z}_t|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}\right) = \tilde{\mathbf{z}}_1'\boldsymbol{\varpi}^{-1}\tilde{\mathbf{z}}_1 + tr\left[\boldsymbol{\varpi}^{-1}(1-\rho^{(0)})^2\sigma_v^{2(0)}(I_M - \sigma_v^{2(0)}V_y^{-1})\right]. \quad (5.5)$$

The last term of Equation (5.1) is equal to

$$E(\mathbf{v}'\mathbf{v}|\mathbf{y}_t, \boldsymbol{\theta}^{(0)}) = \tilde{b}_1'\tilde{b}_1 + tr\left[\sigma_v^{2(0)}\left(I_M - \sigma_v^{2(0)}V_y^{-1}\right)\right]. \quad (5.6)$$

By substituting Equations (5.5) and (5.6) in Equation (5.1), we have

$$\begin{aligned} Q_1 &= E\left(l(B,\sigma_v^2,\rho;y_c)\right) \\ &= C-\frac{1}{2}\log|\varpi|-\frac{1}{2}\left(\tilde{\mathbf{z}}_1'\boldsymbol{\varpi}^{-1}\tilde{\mathbf{z}}_1+tr\left[\boldsymbol{\varpi}^{-1}(1-\rho^{(0)})^2\sigma_v^{2(0)}(I_M-\sigma_v^{2(0)}V_y^{-1})\right]\right) \\ &-\frac{1}{2}\log|\sigma_v^2 I_M|-\frac{1}{2\sigma_v^2}\left(\tilde{b}_1'\tilde{b}_1+tr\left[\sigma_v^{2(0)}\left(I_M-\sigma_v^{2(0)}V_y^{-1}\right)\right]\right). \end{aligned} \tag{5.7}$$

We obtain estimators by solving the estimating equations as:

$$\begin{aligned} 0 &= \frac{\partial Q_1}{\partial \mathbf{B}} = \frac{\partial\left(\tilde{\mathbf{z}}_1'\boldsymbol{\varpi}^{-1}\tilde{\mathbf{z}}_1\right)}{\partial \mathbf{B}} \\ &= \frac{\partial}{\partial \mathbf{B}}\left\{\left(\mathbf{y}_t-\boldsymbol{\rho}^{(0)}\mathbf{y}_{t-1}-\mathbf{h}(\mathbf{x}_t,\mathbf{B})+\boldsymbol{\rho}^{(0)}\mathbf{h}(\mathbf{x}_{t-1},\mathbf{B})-(1-\boldsymbol{\rho}^{(0)})\tilde{b}_1-\beta_1 cos(2\pi ft)-\beta_2 sin(2\pi ft)\right)'\boldsymbol{\varpi}^{-1}\right. \\ &\times\left.\left(\mathbf{y}_t-\boldsymbol{\rho}^{(0)}\mathbf{y}_{t-1}-\mathbf{h}(\mathbf{x}_t,\mathbf{B})+\boldsymbol{\rho}^{(0)}\mathbf{h}(\mathbf{x}_{t-1},\mathbf{B})-(1-\boldsymbol{\rho}^{(0)})\tilde{b}_1-\beta_1 cos(2\pi ft)-\beta_2 sin(2\pi ft)\right)\right\}, \end{aligned} \tag{5.8}$$

$$0=\frac{\partial Q_1}{\partial \boldsymbol{\rho}}=-\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\rho}}\{\log|\varpi|\}-\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\rho}}\left\{\tilde{\mathbf{z}}_1'\boldsymbol{\varpi}^{-1}\tilde{\mathbf{z}}_1+(1-\rho)^2 tr(\boldsymbol{\varpi}^{-1}Var(\mathbf{v}|\mathbf{y}_t,\boldsymbol{\theta}^{(0)}))\right\}, \tag{5.9}$$

where $\varpi = S+\mathcal{S} = S+\Sigma+\rho\Sigma\rho'$ and

$$\tilde{z_1} = \mathbf{y}_t-\boldsymbol{\rho}\mathbf{y}_{t-1}-\mathbf{h}(\mathbf{x}_t,\mathbf{B}^{(0)})+\boldsymbol{\rho}\mathbf{h}(\mathbf{x}_{t-1},\mathbf{B}^{(0)})-(1-\boldsymbol{\rho})\tilde{b}_1-\beta_1 cos(2\pi ft)-\beta_2 sin(2\pi ft).$$

$$\hat{\beta}_1=(cos(2\pi ft))^{-2}\varpi^{-1}\left(cos(2\pi ft)\left(\mathbf{y}_t-\boldsymbol{\rho}\mathbf{y}_{t-1}-\mathbf{h}(\mathbf{x}_t,\mathbf{B}^{(0)})+\boldsymbol{\rho}\mathbf{h}(\mathbf{x}_{t-1},\mathbf{B}^{(0)})-(1-\boldsymbol{\rho})\tilde{b}_1-\beta_2 sin(2\pi ft)\right)\right), \tag{5.10}$$

and

$$\hat{\beta}_2=(sin(2\pi ft))^{-2}\varpi^{-1}\left(sin(2\pi ft)\left(\mathbf{y}_t-\boldsymbol{\rho}\mathbf{y}_{t-1}-\mathbf{h}(\mathbf{x}_t,\mathbf{B}^{(0)})+\boldsymbol{\rho}\mathbf{h}(\mathbf{x}_{t-1},\mathbf{B}^{(0)})-(1-\boldsymbol{\rho})\tilde{b}_1-\beta_1 cos(2\pi ft)\right)\right). \tag{5.11}$$

Also,

$$\hat{\sigma}_v^{2(1)}=\frac{1}{m}\left(\tilde{b}_1'\tilde{b}_1+tr\left[Var(\mathbf{v}|\mathbf{y}_t,\boldsymbol{\theta}^{(0)}))\right]\right). \tag{5.12}$$

Thus the EM algorithm is described as follows:

Step 0: Set r=0 and choose starting value $\boldsymbol{\theta}^{(0)}$.

Step 1: For $r \geq 0$, calculate

$$\tilde{b}_1^{(r+1)}=E\left(\mathbf{v}|\mathbf{y}_t,\boldsymbol{\theta}^{(r)}\right)=\sigma_v^{2(r)}V_y^{-1}(\mathbf{y}_t-\boldsymbol{\mu}_y^{(r)}),$$

and

$$E(\mathbf{v}'\mathbf{v}|\mathbf{y}_t,\boldsymbol{\theta}^{(r)})=\tilde{b}_1'\tilde{b}_1+tr\left[\sigma_v^{2(r)}\left(I_M-\sigma_v^{2(r)}V_y^{-1}\right)\right].$$

Step 2: For $r \geq 0$, compute $\hat{\boldsymbol{\theta}}^{(r+1)}$ based on the Equations (5.8- 5.12).

Step 3. Iterate Steps 1 and 2 from $r = 1$ until reaching convergence. □