International Academic Institute for Science and Technology

International Academic Journal of Social Sciences
Vol.10, No.2, 2023, pp.25-38.

ISSN 2454-3918

International Academic Journal of Social Sciences

www.iaiest.com

Bayesian and Classical Semi-parametric Estimation of the Balanced Longitudinal Data Model

Ahmed Abdulsamad Habeeb [a], Qutaiba N. Nayef Al-Kazaz [b]

[a] *Statistics Department, College of Administration & Economics, University of Baghdad, Baghdad, Iraq.*
E-mail: statisticsahmed@gmail.com
[b] *Statistics Department, College of Administration & Economics, University of Baghdad, Baghdad, Iraq.*
E-mail: dr.qutaiba@coadec.uobaghdad.edu.iq

Abstract

The primary objective of this study is to employ semi-parametric regression techniques in the balanced longitudinal data model. Where the parametric regression models are plagued by the problem of strict constraints, while non-parametric regression models, despite their flexibility, suffer from the problem of the curse of dimensionality. Consequently, semi-parametric regression presents a suitable solution to address the problems in parametric and non-parametric regression methods. The advantage of this model is that it contains all the positive properties included in the previous two models such as containing strict restrictions in its parametric component, complete flexibility in its non-parametric component, and clarity of the interaction between its parametric and non-parametric components.

According to the above, two methods were used to estimate a semi-parametric balanced longitudinal data model. The first is the Bayesian estimating method; the second is the Speckman method, which estimated the unknown nonparametric smoothing function by employing the kernel smoothing Nadaraya & Watson method. The Aim was to make a comparison between the Bayesian estimation method and the classical estimation method. Based on simulation experiments conducted on three different sample sizes (50, 100, and 200), it was concluded that the Bayes method is best at the variance levels (1,5). In contrast, the Profile least square method is best at the variance level (10).

Keywords: Bayesian Method, Profile Least Square Method, Nadaraya &Watson, Semi-parametric Estimation, Balanced Longitudinal Data.

DOI:10.9756/IAJSS/V10I2/IAJSS1010
Received:19 September 2023/Accepted: 18 October 2023/Published: 02 November 2023

Introduction

Longitudinal data can be defined as the data that can be obtained through repeated observations of a phenomenon around (n) cross-sections. Which may be countries, institutions, companies, cities, individuals... etc., and during a certain time (T). (**Zeger et al., 2002**)

Longitudinal data has received great attention in academic studies because it takes into account the effect of time change and the effect of changing cross-sections for the observations.

Longitudinal data refers to a type of data in which the phenomenon under study exhibits changes on two levels: The change in the Vertical level Represents time-series data, While the change on the horizontal level represents the cross-sections data.

Noted that the term longitudinal data gives the same meaning as panel data in the literature of researchers, as for this research it uses the term longitudinal data. Typically, longitudinal data is represented using specialized models, namely regression models designed for longitudinal data, these models have depended heavily on parametric regression models, which are the most commonly used type. Nevertheless, certain applied aspects have raised concerns regarding the suitability of this particular type to represent the phenomenon being studied accurately. This is primarily due to the presence of variables exhibiting parametric behavior while others non-parametric, As well as the failure to consider the nonlinear impact of explanatory variables on the response variable. Hence, the estimations obtained from this type models may potentially lead to misleading conclusions.

There is a second type is the non-parametric regression model that takes into account the nonlinear effect of variables, which is characterized by their high flexibility and depend on the smoothing of the data using weight functions, The most famous of which is the kernel function that is used to smooth the data. However, most researchers have noticed that non-parametric regression also suffers from the curse of dimensionality, which occurs when the number of variables increases. The third type is semi-parametric regression, The great advantage of this model is that it contains all the positive features included in the previous two types and the clarity of the interaction between its parametric and non-parametric components. **(Su, 2011) (Green, 2002)**

The emergence of the semi-parametric approach in regression models is the result of the complement between both parametric and non-parametric inputs because its idea comes from the idea of additive models where the parametric and non-parametric components are combined in this model, Consequently, this type has gained wide acceptance in economic and social studies and other modern studies such as longitudinal data.

The areas of use of the semi-parametric type mentioned above have been expanded to be applied to Bayesian theory. Bayes' theory of statistical inference depends on employing prior information about unknown parameters and considering these parameters as random variables; Hence, this information can be formulated as a probability distribution. It is called the prior probability density function (Prior p.d.f.), and this information is obtained from previous data and experiments or the theory that governs this phenomenon. Bayes' theorem also relies on the current sample information represented by the likelihood function of the observations.

Literature Review

Below are some previous studies in the field of Classical and Bayesian estimation for example:

Wang (2014) developed the proposal that was discussed by each of the researchers (Henderson & Ullah) in their research in (2005) by finding estimates for the non-parametric random effects model for the Longitudinal data, through the two-step method. Through the analysis of simulation experiments, the efficiency of the proposed estimator for all prepared samples was reached, using the comparison standard (MSE).

Khalil and Fadam (2016) studied the Mixed-effects conditional logistic regression in longitudinal pollution data. The research demonstrated that conditional logistic regression is a robust evaluation method for environmental studies. It was shown through simulation that mixed-effects conditional logistic regression is more accurate for pollution studies.

Shaker (2016) submitted a dissertation in which she used parametric and semi-parametric Bayesian methods to estimate the reliability of the systems using the Dirichlet process prior and compared them with the reliability estimations of the systems using the classical methods. The results showed the preference of the Bayesian method for a sample size of n=14.

Liu et al (2017) proposed methods for estimating the parameters of a non-parametric model for Longitudinal data, which was considered one of the important modeling options in the effect of the covariates variable that may change dynamically over time using the correlation function, The researchers also proposed a new method that includes the performance of the selected sample was evaluated by conducting simulation studies, and an example of real data was analyzed to illustrate the proposed methods.

Burhan and Hamoud (2018) compared the estimate of the transfer function using the non-parametric method, represented by two methods: positional linear regression, the cubic bootstrap method, and the semi-parametric method, represented by a single-indicator semi-parametric model with the proposed cubic bootstrap, and the study proved that the proposed estimator is the best among the studied estimators.

Castelein et al (2020) developed a general method for selecting heterogeneous variables in non-linear Longitudinal data models such as polynomial logarithms models based on the Bayesian semi-parametric method and Dirichlet process mixture (DPM) and they reached an improvement in performance in the process of selecting variables heterogeneous.

Nayef and Lina (2022) estimated the missing values for the multivariate skew normal distribution function using the K-nearest neighbors Imputation (KNN). After estimating the missing values, the parameters are estimated using Genetic Algorithm (GA). and the Bayesian Approach was also used to estimate the missing values and find the estimates for the parameters. by comparing the two methods the (GA) that is based on the (KNN)algorithm to estimate the missing values proved to be better and more efficient than the Bayesian Approach in terms of the results.

Nayef and Ali (2022) estimated the analysis of stochastic differential equation with long memory, represented by fractional diffusion process, in this paper they suggested a method for a system of stochastic differential equations with long memory, also they use the Bayesian methodology to incorporate the advanced knowledge, the proposed method has been proved to be very accurate.

The Classical Estimation

Local Linear Polynomial Estimation Method

A method works to correct some defects in the kernel estimator. This method depends on the generalization of the kernel estimator to the case of a compatible polynomial at point (x). It is characterized by the ability to adapt to the nature of the studied model, whether fixed with one smoothing parameter or random with a varying smoothing parameter. Local smoothers can also be one of the most suitable smoothers for the different types of kernel functions and bandwidths used in estimation, so they are classified as highly efficient smoothers. (Hardle, 2004) (Hamoud, 2012)

The local linear polynomial estimation of the Longitudinal data model is based on the Taylor series and as shown in the following formula:

$$m(x_{ij}) \approx m_0(x) + m_1(x)(x_{ij} - x_i) + m_2(x)(x_{ij} - x_i)^2 + \cdots + m_p(x)(x_{ij} - x_i)^p \quad (1)$$

By (p+1) of the series derivatives of the function to be estimated $m(x_{ij})$, we get:

$$m^{(1)}(x_{ij}) = m_1(x) + 2m_2(x)(x_{ij} - x_i) + \cdots + pm_p(x)(x_{ij} - x_i)^{p-1}$$
$$m^{(2)}(x_{ij}) = 2m_2(x) + 6m_3(x)(x_{ij} - x_i) + \cdots + p(p-1)m_p(x)(x_{ij} - x_i)^{p-2}$$
$$m^p(x_{ij}) = p!\, m_p(x)$$

Assuming that:

$$m(x_{ij}) = m_0(x) \Rightarrow m(x_{ij}) = \delta_0$$
$$m^{(1)}(x_{ij}) = m_1(x) \Rightarrow m^{(1)}(x_{ij}) = \delta_1$$
$$m^{(2)}(x_{ij}) = 2m_2(x) \Rightarrow \frac{1}{2} m^{(2)}(x_{ij}) = \delta_2$$
$$m^{(p)}(x_{ij}) = p!\, m_p(x) \Rightarrow \frac{1}{p!} m^{(p)}(x_{ij}) = \delta_p$$

Substituting $\delta_0, \delta_1, \delta_2, \ldots, \delta_p$ instead of $m_0(x), m_1(x), m_2(x), \ldots, m_p(x)$ respectively, Then formula (1) becomes as follows:

$$m(x_{ij}) \approx \delta_0 + \delta_1(x_{ij} - x_i) + \delta_2(x_{ij} - x_i)^2 + \cdots + \delta_p(x_{ij} - x_i)^p \quad \text{Or}$$

$$m(x_{ij}) = \sum_{s=0}^{p} \delta_s (x_{ij} - x_i)^s \quad , s = 0,1, \ldots, p \quad (2)$$

By substituting for $m(x_{ij})$ in the following non-parametric Longitudinal data model, then:

$$y_{ij} = \sum_{s=0}^{p} \delta_s (x_{ij} - x_i)^s + \varphi_{ij} \;, i = 1, \dots, n, j = 1, \dots, t \quad , s = 0,1, \dots, p \tag{3}$$

By determining the value of (p), it is possible to find the estimators of the different local linear polynomial regression under the conditions and assumptions set by the Longitudinal data model with random effects, using the Weighted Least Square (WLS) method. The local linear estimator from the first order can be obtained if it is (p = 1) then the Longitudinal data model with random effects defined in formula (3) will be according to the following formula:

$$m(x_{ij}) \approx \delta_0 + \delta_1 (x_{ij} - x_i)$$
$$y_{ij} = \delta_0 + \delta_1 (x_{ij} - x_i) + \varphi_{ij}$$
$$\varphi_{ij} = x_{ij} - \delta_0 - \delta_1 (x_{ij} - x_i) \tag{4}$$

By applying the weighted least squares method, the sum of squares error of the random model is defined as:

$$\sum_{i=1}^{n} \sum_{j=1}^{t} \varphi_{ij}^2 K_h (x_{ij} - x_i) = \sum_{i=1}^{n} \sum_{j=1}^{t} \{y_{ij} - \delta_0 - \delta_1 (x_{ij} - x_i)\}^2 K_h (x_{ij} - x_i) \tag{5}$$

The above formula can be written in matrices formula as follows:

$$\varphi' K \varphi = \{Y - D'\delta\}' K \{Y - D'\delta\} \tag{6}$$

$$\varphi = \begin{bmatrix} \varphi_1' \\ \varphi_2' \\ \vdots \\ \varphi_n' \end{bmatrix}_{nt\times 1} , \quad Y = \begin{bmatrix} y_1' \\ y_2' \\ \vdots \\ y_n' \end{bmatrix}_{nt\times 1} , D = \begin{bmatrix} 1 & (x_{11} - x)' \\ 1 & (x_{22} - x)' \\ \vdots & \\ 1 & (x_{nt} - x)' \end{bmatrix}_{(p+1)\times nt}$$

$$\delta = \begin{bmatrix} \delta_0 \\ \delta_1 \end{bmatrix}_{(p+1)\times 1} , K = \begin{bmatrix} K_h(x_{11} - x) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & K_h(x_{nt} - x) \end{bmatrix}_{nt\times nt}$$

$$diag(K) = \left[K_h (x_{ij} - x)', \dots, K_h (x_{nt} - x)'\right]$$

To find an estimate of the vector δ, the sum of the squares of the error is reduced by deriving the formula (6) with respect to δ and equating the derivative to zero. We get the following local linear estimator for the first-order random effects Longitudinal data model :(Poo, 2017)

$$\frac{\partial \varphi' K \varphi}{\partial \delta} = -DK[-D\delta] = 0$$
$$\widehat{m}(x_{ij})_{LLLS} = \hat{\delta} = I'(D'KD)^{-1} D'KY \tag{7 }$$

I: It represents a vector with dimensions [(1+p)×1], the first (p) of its elements is equal to 1 and the rest of its elements are zero.

Semi-parametric Regression

Semi-parametric regression is defined as a model containing two components: a parametric component with finite dimensions and a non-parametric with infinite dimensions. There are two approaches to semi-parametric estimation. The first is to estimate the parametric component by any parametric estimation method in the first step. Then, in the second step, the non-parametric part is estimated by any non-parametric estimation method, depending on the estimates of the first step. The second method is opposite to the first method, as the non-parametric component is estimated in the first step, and in the second, the parameter component is estimated based on the estimates of the first step.

In this paper, the method of the Local Linear Polynomial Estimation mentioned in equation (7) was relied upon to estimate the non-parametric part. In contrast, the ordinary least squares method was used to estimate the parametric part.

Partially Linear Longitudinal Data Model with Random Effect

The Partially Linear Model (PLM) is one of the semi-parametric regression models, this model relies on a combination of linear (parametric) and non-linear (non-parametric) variables, which are usually continuous. (Su, 2011) (Green, 2002), The model in question is known by various names, such as the semi-parametric regression model (SPRM) or the partially linear model (PLM).

The semi-parametric partially linear model of longitudinal data is described by the equation below: (Yun,2012) (poo,2017)

$$y_{ij} = x'_{ij}\beta + m(z_{ij}) + \varphi_{ij} \ , i = 1, \dots, n \ \ , j = 1, \dots, t \qquad (8)$$

Where y_{ij}, response variable vector with dimensional nt x 1, x_{ij} Matrix of parametric explanatory variables with dimensional nt x q, β The vector of the unknown parameters with dimensional p x1, $x'B$ The parametric part of the model under study, z_{it} It represents the nonparametric variable in the data with n x 1,$m(z_{it})$: The nonparametric part function is smoothing function with n x1, φ_{it} random error vector.

The semi-parametric linear partially model defined in formula (8) is subject to the following conditions and assumptions:

$E(v_{ij}|x_{ij}, z_{ij}) = 0$
$E(v_{ij}v'_{ij}|x_{ij}, z_{ij}, u_i) = \sigma^2_{ij}I_t$
$E(u_i|x_{ij}, z_{ij}, u_i) = 0$
$E(u_i^2|x_{ij}, z_{ij}, u_i) = \sigma^2_u$
$E(\varphi_{ij}|z_{ij}) = 0$

Hence, the model in (8), according to the above assumptions, will be as follows:

$$E(y_{ij}|x_{ij}, z_{ij}) = x'_{ij}\beta + m(z_{ij}) \qquad (9)$$

Estimation Methods of the Model Semi-Parametric for Longitudinal Data with Random Effect

In the following section (2.3.1), We will review the semi-parametric method used in the research to estimate the semi-parametric model of the longitudinal data in equation (8) as:

Profile Least Square Method

This method was proposed in 1988 by Robinson for estimating the partially linear model based on converting the semi-parametric model into a non-parametric model by subtracting the parametric part of the semi-parametric model. **(**Robinson, 1988),

Then (Li & Stengos, 1996) developed this method to fit the Longitudinal data, and it was named by (Fan & Huang, 2005) as (Profile Least Square Method) (PLLLS). (Fan, 2005) (Poo, 2017) (Su, 2011) (Yun, 2012).

This method can be applied to the partially linear model of Longitudinal data in the following steps:

$$Y_{ij} - X^T_{ij}\beta = m(Z_{ij}) + \varphi_{ij} \quad , i = 1, \dots, n, \qquad j = 1, \dots, t$$

$$Y^*_{ij} = Y_{ij} - X^T_{ij}\beta$$

$$Y^*_{ij} = m(Z_{ij}) + \varphi_{ij} \qquad (10)$$

It is possible to estimate the unknown smoothing function using the local polynomial linear estimation according to the methods used in the non-parametric estimation: Assuming that θ_0 and θ_1 represent the resulting solutions after minimizing the sum of squares of the error of the semi-parametric model as follows: (poo,2017)

$$\sum_{i=1}^{n}\sum_{j=1}^{t}\left[(y_{ij} - x'_{ij}\beta) - \theta_0 - \theta_1(z_{ij} - z_i)\right]^2 K_h(z_{ij} - z_i) \qquad (11)$$

Note that $\hat{\theta}' = (\hat{\theta}_0, \hat{\theta}_1)$ represents estimates for $m_0(z_{ij})$ and $m_1(z_{ij})$, respectively, so that:

$$\hat{\theta} = I_{NT}(H'K_zH)^{-1}H'K_z(Y - Z\beta) = S(Y - Z\beta) \qquad (12)$$

Where I_{nt} represents q vector, the first p from its element is one and the remainder is zero, S smoothing matrix, $S = I_{nt}(H'K_zH)^{-1}H'K_z$, K The matrix of weights for the diagonal kernel function with dimensions (nt x nt):

$diag(K) = [K_h(Z_{11} - Z)', \dots, K_h(Z_{nt} - Z)']$, Y response variable vector with dimensions (nt x 1), $Y = [Y_{11}, \dots, Y_{nt}]$. X the matrix of explanatory variables with dimensions (nt x q) , $X = [X_{11}, \dots, X_{nt}]$, H matrix with dimensions (nt) x (1+p) , in which

$$H = \begin{bmatrix} 1 & \cdots & (z_{11} - z)' \\ \vdots & \ddots & \vdots \\ 1 & \cdots & (z_{nt} - z)' \end{bmatrix}.$$

By substituting $m(z_{ij})$ into formula (10), the semi-parametric regression model for Longitudinal data can be written in Matrices and vectors form as follows:

$$\hat{Y} = \hat{X}'\beta + w^* \qquad i = 1, \dots, n, \qquad t = 1, \dots ., t \tag{13}$$

So that:

$\hat{Y} = (I_{nt} - S)Y \quad , \hat{Y} = (\hat{Y}_{11}, \dots, \hat{Y}_{nt})$

$\hat{X} = (I_{nt} - S)X \quad , \hat{X} = (\hat{X}_{11}, \dots, \hat{X}_{nt})$

$w^* = (I_{nt} - S)w_i + (I_{nt} - S)m(z) \quad , w^* = (w^*_{11}, \dots ., w^*_{nt})$

$m(z) = m(z_{11}), \dots ., m(z_{nt})$

When applying the least squares method, the semi-parametric estimation of the parameters vector (β) and the smoothing function $m(z)$ is obtained, as shown below:(poo,2017)

$$\hat{\beta}_{PLLLS} = (\hat{X}'\hat{X})^{-1}(\hat{X}'\hat{Y}) \tag{14}$$

$$\widehat{m}_{PLLLS}(z) = I_{NT}(H'KH)^{-1}H'K\left(Y - X\hat{\beta}_{PLLLS}\right) \tag{15}$$

The Bayesian Estimation

Posterior Distributions

It is defined as a function that represents all the information about the parameters to be estimated after observing the sample information. It is also called distribution after sampling, The notation denotes $\pi(\theta/D)$ the posterior distribution is a conditional probability distribution for the parameter β with the condition that sample x is obtained and assuming that θ is a random variable that has a prior distribution and denoted by the notation $\pi(\theta)$, The inference of θ is based on the posterior distribution, which we obtain by Bayes' theorem, The following formula gives the posterior distribution of the random variable θ:

$$\pi(\theta/D) = \frac{L(\theta/D)\,\pi(\theta)}{\int_\Theta L(\theta/D)\,\pi(\theta)d\theta} \tag{16}$$

Where: Θ represent the parameter range θ, $L(\theta/D)$ Likelihood function

From the above formula, it is clear that $\pi(\theta/D)$ is proportional to the Likelihood function multiplied by the prior distribution.

$\pi(\theta/D) \propto L(\theta/D)\,\pi(\theta)$

So, it includes data contribution by $L(\theta/D)$, And the contribution of the primary information specified by $\pi(\theta)$.

Dirichlet Processes

The Diricihlet process is a family of random processes whose products are probability distributions. In other words, it is a probability distribution, which in itself is a set of probability distributions. It is often used in Bayesian inference to describe prior knowledge about the distribution of random variables, that is, the extent likely that the random variables will be distributed according to one particular distribution. The Dirichlet process is specified by the base distribution, which is denoted by (H) and a positive real number α called the scaling parameter (also known as concentration parameter). The base distribution is the expected value of the process, which means the Dirichlet process draws distributions "around" the base distribution like a normal distribution draws real numbers around its mean. However, even if the base distribution is continuous, the distributions drawn from the Dirichlet process are almost surely discrete, the (α) specifies how strong this discretization is in the limit of α →0, the realizations are all concentrated at a single value, while in the limit of α→∞ the realizations become continuous. Between the two extremes, the realizations are discrete distributions with less and less concentration as α increases.

It is worth noting that the Dirichlet process is used to find the prior distributions of non-parametric functions, since then, it has been applied in the fields of data mining, machine learning, arithmetic, and counting, as well as in data science and information. (Ferguson, 1973)

To illustrate the Dirichlet process, assume that Θ is a random distribution that will be distributed according to Dirichlet, and we assume that H is the base distribution and α is a positive real number.

$\Theta \sim DP(\alpha, H)$

The vector $\Theta = \{\theta_1, \theta_2, \dots, \theta_m\}$ is random, where we say that G is distributed according to the Dirichlet process with the base distribution H and the concentration (scale) parameter α such that it is written in the following form: (Ferguson, 1973)

$\Theta \sim Dirichlet(\alpha_1, \alpha_2, \dots, \alpha_m)$

A Semiparametric Mixed Models

In order to achieve flexibility in applying the Bayesian estimation, the semi-parametric linear partially model in formula (8) can be converted into a semi-parametric mixed model as in formula (17) below. (AL-Mouel & Mohaisen, 2017)

Where the random effects model or the semi-parametric mixed model in Longitudinal data models, random effects are usually assumed to be normally distributed in both Bayesian and classical models. Depending on the Bayesian approach, it can allow random effects to obtain a prior non-parametric distribution, employing the Dirichlet process. The calculation of this thing became possible by Gibbs Sampler. The general form of the semi-parametric mixed model can be defined for the Longitudinal data as follows: (Kleinman& Ibrahim, 1998) (Laird & Ware, 1982)

$$y_{it} = x'_{it}\beta + w'_{it}b_i + \varepsilon_{it}\,, \qquad i = 1, \dots, n, t = 1,,,, T. \tag{17}$$

Where: y_{it} Represents the response variable, which is an n×1 vector, x_{it} is an n×p fixed explanatory variables matrix, β the p×1 dimensional fixed effects vector of the regression coefficients, which are commonly referred to as fixed effects in these models, w_{it} is an n×q random explanatory variables matrix, bi A random effects vector with dimensions q×1, ε_{it} vector with dimension n×1 of errors. And $\varepsilon_{it} = v_i + u_{it}$

Assuming that ε_{it} and bi are independent and each of them is normally distributed by:

$$\varepsilon_{it} \sim N(0, \tau^{-1}) \tag{18}$$
$$b_i \sim N(0, D) \tag{19}$$

Bayesian Semiparametric Estimation

Finding the Estimation of the Parametric Component

If the distribution of observations y, a normal distribution, is described as follows:

$N(x'_{it}\beta + w'_{it}b_i, \tau)$

Then Likelihood function: (Jochmann & León, 2004).

$$\prod_{i=1}^{n} f(y \backslash x, \beta, \tau) = \left(\frac{1}{\sqrt{2\pi\tau}}\right)^{n} \times exp\left(\frac{-\tau}{2}\sum_{i=1}^{n}\sum_{t=1}^{T_i}(y_{it} - x'_{it}\beta - w'_{it}b_i)^2\right) \tag{20}$$

And the prior distribution for β is normal distribution $N(\mu_0, \Sigma_0)$

$$\pi_0(\beta) = |\Sigma_0|^{-1/2} exp\left(\frac{-1}{2}(\beta - \mu_0)'\Sigma_0^{-1}(\beta - \mu_0)\right) \tag{21}$$

We will find the posterior distribution as follow: (Jochmann & León, 2004).

$$\begin{aligned} &p(\beta|\{b_i\}, \tau, \{y_{it}\}) \\ &\propto |\Sigma_0|^{-1/2} exp\left(\frac{-1}{2}(\beta - \mu_0)'\Sigma_0^{-1}(\beta - \mu_0)\right) \\ &\times exp\left(\frac{-\tau}{2}\sum_{i=1}^{n}\sum_{t=1}^{T_i}(y_{it} - x'_{it}\beta - w'_{it}b_i)^2\right) \\ &= exp\left\{\frac{-\tau}{2}\sum_{i=1}^{n}\sum_{t=1}^{T_i}(y_{it} - x'_{it}\beta - w'_{it}b_i)^2 + \left(\frac{-1}{2}(\beta - \mu_0)'\Sigma_0^{-1}(\beta - \mu_0)\right)\right\} \\ &= exp\left\{\frac{-\tau}{2}\sum_{i=1}^{n}\sum_{t=1}^{T_i}\begin{pmatrix} y_{it}^2 + \beta' x' x\beta + b' w w' b + 2b' w x'\beta - 2 \\ y x'_{it}\beta - 2y w'_{it} b_i \end{pmatrix} + \frac{-1}{2}(\beta'\beta + \mu'_0\mu_0 - 2\beta\mu_0)'\Sigma_0^{-1}\right\} \end{aligned} \tag{22}$$

Then, comparing the resulting distribution with the normal distribution to obtain the covariance of the resulting distribution:

$$exp\left\{\frac{-1}{2\sigma^2}[\,\mu^2 + \mu_n^2 - 2\mu\mu_n\,]\right\}$$

$$\frac{1}{2\sigma_n^2}\beta'\beta = \frac{\beta'\beta}{2}[\tau\sum_{i=1}^{n}\sum_{t=1}^{T} x' x + \Sigma_0^{-1}]$$

$$\frac{1}{\sigma_n^2} = [\tau \sum_{i=1}^{n} \sum_{t=1}^{T} x'x + \Sigma_0^{-1}]$$

$$\sigma_n^2 = \left[\tau \sum_{i=1}^{n} \sum_{t=1}^{T} x'x + \Sigma_0^{-1}\right]^{-1}$$

$$\Sigma_\beta = \left[\tau \sum_{i=1}^{n} \sum_{t=1}^{T} x'x + \Sigma_0^{-1}\right]^{-1} \quad (23)$$

To obtain the mean:

$$\frac{-1}{2}(-2\beta'\mu_0)\Sigma_0^{-1} + \frac{-\tau}{2} \sum_{i=1}^{n} \sum_{t=1}^{T} (2w'b\beta'x - 2y\beta'x)$$

$$\frac{-1}{2}(-2)(\beta'\mu_0)\Sigma_0^{-1} + \frac{-\tau}{2} 2\beta' \sum_{i=1}^{n} \sum_{t=1}^{T} x(w'b - y)$$

$$(\beta'\mu_0)\Sigma_0^{-1} - \tau\beta' \sum_{i=1}^{n} \sum_{t=1}^{T} x(w'b - y)$$

$$\beta'\left[\mu_0\Sigma_0^{-1} + \tau \sum_{i=1}^{n} \sum_{t=1}^{T} x(y - w'b)\right]$$

$$\frac{2\mu\beta'}{2\sigma_n^2} = \beta'\left[\mu_0\Sigma_0^{-1} + \tau \sum_{i=1}^{n} \sum_{t=1}^{T} x(y - w'b)\right]$$

$$\frac{\mu}{\sigma_n^2} = \left[\mu_0\Sigma_0^{-1} + \tau \sum_{i=1}^{n} \sum_{t=1}^{T} x(y - w'b)\right]$$

Then the Bayes estimator:

$$\mu_\beta = \Sigma_\beta \left[\mu_0\Sigma_0^{-1} + \tau \sum_{i=1}^{n} \sum_{t=1}^{T} x(y - w'b)\right] \quad (24)$$

Then, find an estimate of the variance: if the variance is unknown,

$\tau = \sigma^{-2}$

If τ follows the following gamma distribution:

$$\tau \sim Gamma\ \left(\frac{\alpha_0}{2}, \frac{\alpha_0}{2}\right) \quad (25)$$

$$\pi(\tau) = \frac{\left(\frac{\alpha_0}{2}\right)^{\frac{\alpha_0}{2}}}{\Gamma\left(\frac{\alpha_0}{2}\right)} \tau^{\frac{\alpha_0}{2}-1} exp\ \left\{\frac{-\alpha_0}{2}\tau\right\} \quad (26)$$

$$p(\tau|\{b_i\}, \beta, \{\theta_{it}\}) \propto\ \tau^{(\alpha_0/2)-1} exp\left(\frac{-\alpha_0\tau}{2}\right)$$

$$\times\ \tau^{n/2} exp\left(\frac{-\tau}{2} \sum_{i=1}^{n} \sum_{t=1}^{T_i} \varepsilon_{it}^2\right)$$

$$\propto \tau^{(\alpha_0+n/2)-1} exp\ \left\{\frac{-\tau}{2}\left[\alpha_0 + \sum_{i=1}^{n} \sum_{t=1}^{T} \varepsilon_{it}^2\right]\right\}$$

Which results in a gamma distribution with the following parameters: (Jochmann & León, 2004)

$$[\tau|\{b_i\}, \beta, \{\theta_{it}\}] \sim Gamma\left(\frac{\alpha_0 + n}{2}, \frac{\alpha_0 + \sum_{i=1}^{n}\sum_{t=1}^{T_i} \varepsilon_{it}^2}{2}\right) \quad (27)$$

Finding the Estimation of the Nonparametric Component Based on the Dirichlet Distribution

Semi-parametric mixed model in Longitudinal data models, the random effects of the nonparametric component are usually assumed to be normally distributed in Bayesian models. It is possible to depend on the Bayesian approach to allow random effects to obtain a prior non-parametric distribution, depending on the Dirichlet process.

The posterior conditional distribution of bi can be expressed according to the Dirichlet distribution on b^{-i} and G0 as follows: (Jochmann & León, 2004).

$$b_i \backslash b^{-i}, G_0 \sim \frac{M}{M+N-1} G_0 + \frac{1}{M+N-1} \sum_{j=1}^{l} m_j^{-i} \delta\left(k_j^{-i}\right) \quad (28)$$

So, if the distribution of parameter bi follows the normal distribution as mentioned above in equation (19).

We can now merge this result with the Likelihood function that follows the normal distribution and take the integral to obtain the following result: (Jochmann & León, 2004).

$$b_i \backslash \beta, D, G_0, b^{-i} \sim \int \emptyset(y_{it} \backslash x'_{it}\beta + w'_{it} b_i, \tau^{-1}) d[b_i \backslash b^{-i}, G_0]$$

$$b_i \backslash \beta, D, G_0, b^{-i} \sim \int \emptyset(y_{it} \backslash x'_{it}\beta + w'_{it} b_i, \tau^{-1}) \left[\frac{M}{M+N-1} G_0 + \frac{1}{M+N-1} \sum_{j=1}^{l} m_j^{-i} \delta\left(k_j^{-i}\right)\right] \quad (29)$$

We performing the integration we end up with:

$$b_i \backslash \beta, D, G_0, b^{-i} \sim$$
$$\left[\frac{M}{M+N-1} \emptyset(y_{it} \backslash x'_{it}\beta + w'_{it} b_i, \tau^{-1}) + \frac{1}{M+N-1} \sum_{j=1}^{l} m_j^{-i} \emptyset(y_{it} \backslash x'_{it}\beta + w'_{it} b_i, \tau^{-1})\right]$$

$$b_i \backslash \beta, D, G_0, b^{-i} \sim$$
$$\frac{1}{M+N-1} \left[M \emptyset(y_{it} \backslash x'_{it}\beta + w'_{it} b_i, \tau^{-1}) + \sum_{j=1}^{l} m_j^{-i} \emptyset(y_{it} \backslash x'_{it}\beta + w'_{it} b_i, \tau^{-1})\right] \quad (30)$$

The final bi-posterior distribution is obtained by: (west et al., 1994)

$$p(bi \backslash \beta, \tau, y, b_{-i}) \propto$$
$$\left\{M \int \emptyset(y_{it} \backslash X_{it}\beta + w_{it} b_i, \tau^{-1}) \emptyset(b_i \backslash 0, D) db_i \right\} \times \emptyset(b_i \backslash 0, d)\, p\left(y_{it} \backslash b_i, \beta, \tau, y_{jt}\right)$$
$$+ \sum_{j \neq i} \emptyset(\left(y_{it} \backslash X_{it}\beta + w_{it} b_j, \tau^{-1}\right). \delta_{b_j}$$

After several algebraic operations, we get: **(**Kleinman& Ibrahim, 1998)

$$p(b_i \backslash \beta, \tau, y, b_{-i}) \propto M |\Sigma_b|^{1/2} |D|^{-1/2} \tau^{ni/2} \times \emptyset(b_i \backslash 0, D) p(y_{it} \backslash b_i, \beta, \tau)$$
$$+ \left(\sum_{j \neq i} \tau^{ni/2} exp\left[\frac{-\tau}{2}\left(y_{it} - X_{it}\beta - w_{it} b_j\right)'\left(y_{it} - X_{it}\beta - w_{it} b_j\right)\right]. \delta_{b_j}\right)$$

Each term is separated into two elements; the first element is a mixing probability, and the second is a distribution to be mixed, So, the second term is the probability of mixing is proportional to: **(**Kleinman& Ibrahim, 1998)

$$\tau^{ni/2} \exp\left(\frac{-\tau}{2}(y_{it} - X_{it}\beta - w_{it} b_j)'(y_{it} - X_{it}\beta - w_{it} b_j)\right)$$

We select from distribution δ_{b_j}, which means that we set $b_i = b_j$, Also with probability proportional to:

$$M |\Sigma_b|^{1/2} |D|^{-1/2} \tau^{\frac{ni}{2}} \int \exp\left\{\frac{\tau}{2}\left[(y_{it} - X_{it}\beta)' U_i (y_{it} - X_{it}\beta)\right]\right\} db_i$$

We select from

$$p(b_i \backslash \beta, \tau, y_{it}) \propto \emptyset(b_i \backslash 0, D) p\left(y_{it} \backslash b_i, \beta, \tau, y_{ij}\right) \quad (31)$$

Where:

$U_i = (\tau W_i \Sigma_b W'_i - I)$

And Σ_b the covariance matrix of b_i which will be found by deriving the posterior distribution in equation (31) as follows:

$$p(b_i|\beta,\tau,D,\{\theta_{it}\}) \propto exp\left(\frac{-1}{2}b_i'D^{-1}b_i\right)$$
$$\times\, exp\left(\frac{-1}{2}\sum_{t=1}^{T_i}(\theta_{it}-x_{it}'\beta-w_{it}'b_i)^2\right)$$
$$exp\left\{\frac{-\tau}{2}\sum_{i=1}^{n}\sum_{t=1}^{T_i}(y_{it}-x_{it}'\beta-w_{it}'b_i)^2+\left(\frac{-1}{2}b_i'D^{-1}b_i\right)\right\}$$
$$exp\left\{\frac{-\tau}{2}\sum_{i=1}^{n}\sum_{t=1}^{T_i}\left(\begin{matrix}y_{it}^2+\beta'x'x\beta+b'ww'b+2b'wx'\beta-2\\ yx_{it}'\beta-2yw_{it}'b_i\end{matrix}\right)+\left(\frac{-1}{2}b_i'D^{-1}b_i\right)\right\}$$

Then, comparing the resulting distribution with the normal distribution to obtain the covariance of the resulting distribution:

$$exp\left\{\frac{-1}{2\sigma^2}[\mu^2+\mu_n^2-2\mu\mu_n]\right\}$$
$$\frac{b'b}{2\sigma_n^2}=\left[\frac{-1}{2}b'D^{-1}b+\frac{-\tau}{2}\sum_{i=1}^{n}\sum_{t=1}^{T}b'ww'b\right]$$
$$\frac{-b'b}{2\sigma_n^2}=\frac{-b'b}{2}\left[D^{-1}+\tau\sum_{i=1}^{n}\sum_{t=1}^{T}ww'\right]$$
$$\frac{1}{\sigma_n^2}=\left[D^{-1}+\tau\sum_{i=1}^{n}\sum_{t=1}^{T}ww'\right]$$
$$\sigma_n^2=\left[D^{-1}+\tau\sum_{i=1}^{n}\sum_{t=1}^{T}ww'\right]^{-1}$$
$$\Sigma_b=\left[D^{-1}+\tau\sum_{i=1}^{n}\sum_{t=1}^{T}ww'\right]^{-1} \qquad (32)$$

Then, obtain the mean:

$$\frac{2\mu b'}{2\sigma_n^2}=\frac{-\tau}{2}[2b'wx'\beta-2yb'w]$$
$$\frac{\mu b'}{\sigma_n^2}=\frac{-\tau}{2}2b'\,w[x'\beta-y]$$
$$\frac{\mu}{\sigma_n^2}=-\tau\sum_{i=1}^{n}\sum_{t=1}^{T}w[x'\beta-y]$$
$$\mu=\sigma_n^2\left(-\tau\sum_{i=1}^{n}\sum_{t=1}^{T}w[x'\beta-y]\right)$$
$$\mu=\sigma_n^2\left(\tau\sum_{i=1}^{n}\sum_{t=1}^{T}w[y-x'\beta]\right)$$

Then the Bayes estimator is:

$$\mu_b=\Sigma_b\left(\tau\sum_{i=1}^{n}\sum_{t=1}^{T}w[y-x'\beta]\right) \qquad (33)$$

Covariance Estimation

If D is assumed to follow the Wishart distribution

$$D^{-1}\sim Wishart\,(v_0,S_0) \qquad (34)$$
$$f(D^{-1})=\frac{|D^{-1}|^{(v_0-p-1)/2}\,e^{-\frac{-1}{2}tr(S_0^{-1}D^{-1})}}{2^{\frac{v_0p}{2}}|S|^{\frac{v_0}{2}}\,\Gamma p(\frac{v_0}{2})} \qquad (35)$$

Then, after getting rid of the constants and multiplying the prior distribution by the likelihood function of b_i, we get: (west et al., 1994)

$$p(D^{-1}|\{b_i\})$$
$$\propto |D^{-1}|^{(v_0-p-1)/2} exp\ (\frac{-1}{2} tr(S_0^{-1}D^{-1})$$
$$\times |D^{-1}|^{n/2} exp\left(\frac{-1}{2}\sum_{i=1}^{n} b_i' D^{-1} b_i\right)$$
$$\propto |D^{-1}|^{\frac{v_0+n-1}{2}} exp\left\{\frac{-D^{-1}}{2}\left(S_0^{-1}+\sum_{i=1}^{n} b_i' b_i\right)\right\}$$

Therefore, the posterior distribution will be a Wishart distribution with the following parameters:

$$[D^{-1}|\{b_i\}] \sim Wishart\left(v_0+n, \left(S_0^{-1}+\sum_{i=1}^{n} b_i b_i'\right)^{-1}\right) \quad (36)$$

Bayesian MCMC Sampling

A Technique is an essential element of the Monte Carlo Markov Chain (MCMC) methodology and plays a critical role in analyzing the posterior distribution Bayesian estimation. The posterior distribution of the model contains all the information related to the prior distribution and the likelihood function and can be used to provide probability data about the parameters.

However, due to the complexity of the studied models, it isn't easy to analyze their posterior distributions, this problem can be overcome by applying Markov Chain Monte Carlo (MCMC) techniques, where large samples are drawn from the posterior distributions, Then, these samples are used to summarize the posterior distributions. This is done by employing the Gibbs Sampler, where the vector of each parameter is updated by taking its conditional distribution over the rest of the parameters of the other components. After eliminating some of the initial draws, the resulting Markov chains converge for posterior distributions. Sampling continues until the asymptotic posterior distributions are reached. Below is a summary of the Gibbs Sampler algorithm based on posterior distributions:

(Chen et al., 2000) (Robert& Casella, 1999) (Tanner &wong, 1987)

1. Choose starting values for $\tau, \{b_i\}, D^{-1}, \{y_{it}\}$.
2. Sample β from $[\beta|\{b_i\}, \tau, \{y_{it}\}]$, which is a normal distribution.
3. Sample τ from $[\tau|\{b_i\}, \beta, \{y_{it}\}]$, which is a Gamma distribution.
4. Sample y from $[\{y_{it}\}|\{b_i\}, \beta, \tau]$, which is a Normal distribution for i=1,….,N, and t=1,….,T.
5. Sample {bi} from $[b_i|\beta, \tau, D, \{y_{it}\}]$, which is a Normal distribution independently for i=1,….,N.
6. Sample D^{-1} from $[D^{-1}| \{b_i\}]$, which is a Wishart distribution.
7. Repeat Steps 1-5 using the updated values of the conditioning variables.

The Simulation

The R Language program was used to carry out simulation experiments using ten cross sections (n = 10) with three time periods (t = 20, t = 10, t = 5). Thus, we will have three sizes of samples: (NT = 50), (NT = 100), and (NT = 200). Each experiment was repeated for accurate and stable results (Replicates = 1000). The nonparametric smoothing functions that were used in this research are (linear, quadratic, and exponential), respectively, which were taken from published research as follows: (Wang et al., 2004)

$$a)\ m_1(z_{ij}) = z \quad (37)$$
$$b)\ m_2(z_{ij}) = 3.2\ z^2 - 1 \quad (38)$$
$$c)\ m_3(z_{ij}) = exp\ \{(-32(z-0.5)^2)\} \quad (39)$$

The following table (1) describes the models estimated according to the estimation methods.

Table1: Describes the Models Estimated

No.	Model	The Method Estimation
1	Model I	Profile Least Square Method
2	Model II	Bayes Method

Presentation and Analysis of Results

The values of the Average Mean Squared Error (AMSE) for the different models for different levels of variance and different sample sizes using the nonparametric smoothing functions in equations (37,38 and 39) are presented in the following tables:

Table 2: Values of Average Mean Squared Error (AMSE) by using Linear Function in Equation (37)

σ_e^2	Model	AMSE			
$\sigma_e^2 = 1$		**M(z)**	**n=50**	**n=100**	**n=200**
	Model I		0.193344142	0.217206088	0.21911267
	Model II		0.162442607	0.179088110	0.199329073
$\sigma_e^2 = 5$	Model I		2.48216471	2.98690923	3.0368567
	Model II		1.3284889	1.50316086	1.7163896
$\sigma_e^2 = 10$	Model I		9.8345592	10.6007533	11.3865656
	Model II		18.7009065	18.7559739	28.845967

Table 3: Values of Average Mean Squared Error (AMSE) by using Quadratic Function in Equation (38)

σ_e^2	Model	AMSE			
		M(z)	**n=50**	**n=100**	**n=200**
$\sigma_e^2 = 1$	Model I		0.170492711	0.186854371	0.21372524
	Model II		0.152887581	0.17447287	0.196816504
$\sigma_e^2 = 5$	Model I		2.2515057	2.70475752	2.95388926
	Model II		1.3951073	1.4082165	1.452309144
$\sigma_e^2 = 10$	Model I		9.30315377	10.192994	10.8142248
	Model II		10.548904	16.3183375	23.6989039

Table4: Values of Average Mean Squared Error (AMSE) by using Exponential Function in Equation (39)

σ_e^2	Model	AMSE			
		M(z)	**n=50**	**n=100**	**n=200**
$\sigma_e^2 = 1$	Model I		0.124676666	0.135574008	0.165956888
	Model II		0.11476246	0.120329479	0.157833968
$\sigma_e^2 = 5$	Model I		2.21673583	2.2471337	2.40829918
	Model II		0.99439412	1.17621229	1.34821452
$\sigma_e^2 = 10$	Model I		8.899326	10.1610638	10.2708656
	Model II		10.3424103	10.6181225	13.19540385

The values of Average Mean Squares Error in Tables (2,3,4) showed the following results:

1- The models estimated according to the Bayes method (Model II) gave average mean square error values at the variance levels (1,5) less than those provided by the Profile least square method (Model I).
2- At variance levels (10), the models estimated using the Profile least square method (Model I) yielded average mean square error values lower than those obtained using the Bayes method (Model II).
3- The values of the average mean square error estimated increase when the level of variance for the two methods increase.
4- The average mean square error values increase with increasing sample size.
5- Increasing the level of variance leads to the Profile least square method being better than the Bayes method.

Conclusions

After the simulation experiments were carried out and the results presented and analyzed, the researcher concluded the following:

1- Increasing the variance level increases the average mean square error values. Therefore, the relationship between the variance level and the efficiency of the estimation methods is inverse.
2- The Bayes method was best at low variance levels, While the Profile least square method was best at high variance levels.
3- Increasing the sample size leads to an increase in the mean square error of the two methods, which means that the relationship between sample size and the efficiency of the mean square error is inverse.
4- Using the high variances makes it easy for us to determine which method is more efficient compared to the other.

Recommendations

Drawing on the researcher's conclusions, the following recommendations can be summed up to complement and align with the study's concept and subject matter:

1- Should look for other, more efficient methods that are good at low and high variances.
2- Need to Estimating the Longitudinal data model in the case of fixed effects with the Bayesian approach.
3- The semi-parametric partially linear model should be used for Longitudinal data in case of polluting random errors like the presence of outlier data or measurement errors.
4- Expanding the use of Longitudinal data models that are not used in this research, such as the single index model.

References

AL-Mouel, A.S., Mohaisen, A.J., Ionescu, L.M., & Li, Y. (2017). Study on Bayes semiparametric regression. *Journal of Advances in Applied Mathematics, 2*(4), 197-207.

Manaf Youssef Hammoud, & Yaqeen Khalil Burhan. (2018). comparison between the methods estimate nonparametric and semiparametric transfer function model in time series the Using simulation. *journal of Economics and Administrative Sciences, 24* (106), 375-391.

Castelein, A., Fok, D., & Paap, R. (2020). Heterogeneous Variable Selection in Nonlinear Panel Data Models: A Semi-Parametric Bayesian Approach.

Chen, M.H., Shao, Q.M., & Ibrahim, J.G. (2012). *Monte Carlo methods in Bayesian computation*. Springer Science & Business Media.

Fan, J., & Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, *11*(6), 1031-1057.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The annals of statistics*, 209-230.

Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.

Hammoud, M.Y., Ashour, M.A.H. (2012). Comparing a few nonlinear estimators to estimate the regression function. *Journal of Administrative and Economic Sciences, 18*(68).

Härdle, W., Müller, M., Sperlich, S., & Werwatz, A. (2004). *Nonparametric and semiparametric models* (Vol. 1). Berlin: Springer.

Jochmann, M., & León-González, R. (2004). Estimating the demand for health care with panel data: a semiparametric Bayesian approach. *Health Economics*, *13*(10), 1003-1014.

Khalil, Y. & Fadam, E. (2016). Compare to the conditional logistic regression models with fixed and mixed effects for longitudinal data. *Journal of Economics and Administrative Sciences, 23*(98), 406-429.

Kleinman, K.P., & Ibrahim, J.G. (1998). A semiparametric Bayesian approach to the random effects model. *Biometrics*, 921-938.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 963-974.

Li, Q., & Stengos, T. (1996). Semiparametric estimation of partially linear panel data models. *Journal of econometrics*, *71*(1-2), 389-397.

Liu, S., You, J., & Lian, H. (2017). Estimation and model identification of longitudinal data time-varying nonparametric models. *Journal of Multivariate Analysis*, *156*, 116-136.

Nauef Al-Qazaz, Q.N., & Ali, A.H. (2022). Bayesian inference of fractional brownian motion of multivariate stochastic differential equations. *International Journal of Nonlinear Analysis and Applications*, *13*(1), 2425-2454.

Nayef Al-Qazaz, Q.N., & Shawkat, L.N. (2022). Using some methods to estimate the parameters of the Multivariate Skew Normal (MSN) distribution function with missing data. *International Journal of Nonlinear Analysis and Applications*, *13*(1), 2333-2350.

Robert, C.P., Casella, G., & Casella, G. (1999). *Monte Carlo statistical methods*, 2. New York: Springer.

Robinson, P.M. (1988). Root-N-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, 931-954.

Rodriguez-Poo, J. M., & Soberon, A. (2017). Nonparametric and semiparametric panel data models: Recent developments. *Journal of Economic Surveys*, *31*(4), 923-960.

Qutaiba Nabil Nayef, & Aseel Mahmoud Shaker. (2017). Comparison Semiparametric Bayesian Method with Classical Method for Estimating Systems Reliability using Simulation Procedure. *journal of Economics And Administrative Sciences , 23* (97).

Su, L., & Ullah, A. (2011). Nonparametric and semiparametric panel econometric models: estimation and testing. *Handbook of empirical economics and finance*, 455-497.

Tanner, M.A., & Wong, W.H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, *82*(398), 528-540.

Wang, H. (2014). *Essays on Semiparametric Ridge-Type Shrinkage Estimation, Model Averaging and Nonparametric Panel Data Model Estimation*. University of California, Riverside.

West, M., & Escobar, M.D. (1993). *Hierarchical priors and mixture models, with application in regression and density estimation*. Institute of Statistics and Decision Sciences, Duke University.

Wang, Y. (2012). *Essays on Nonparametric and Semiparametric Models and Continuous Time Models*. University of California, Riverside.

Diggle, P. (2002). *Analysis of longitudinal data*. Oxford university press.