

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2022; 7(2): 01-13
© 2022 Stats & Maths
www.mathsjournal.com
Received: 02-01-2022
Accepted: 05-02-2022

Kung'u John
Mathematics and Actuarial
Science Department,
Kenyatta University,
P.O Box: 43844-00100,
Nairobi, Kenya

Kube Ananda
Mathematics and Actuarial
Science Department,
Kenyatta University,
P.O Box: 43844-00100,
Nairobi, Kenya

Odongo Leo
Mathematics and Actuarial
Science Department,
Kenyatta University,
P.O Box: 43844-00100,
Nairobi, Kenya

Corresponding Author:
Kung'u John
Mathematics and Actuarial
Science Department,
Kenyatta University,
P.O Box: 43844-00100,
Nairobi, Kenya

Bayesian approach to zero-inflated dynamic panel ordered probit model with an application in drug abuse

Kung'u John, Kube Ananda and Odongo Leo

DOI: <https://doi.org/10.22271/math.2022.v7.i2a.787>

Abstract

The Zero inflated ordered categorical data with time series structure are often a characteristic of behavioural research attributed to non-participation decision and zero consumption of substance such as drugs among the participants. The existing Semi-parametric zero inflated dynamic panel probit model with selectivity have exhibited biasness and inconsistency in estimators as a result of poor treatment of initial condition and exclusion of selectivity in the unobserved individual effects respectively. The model assumed that the cut points are known to address heaping in the data and therefore cannot be used when the cut points are unknown. In this paper, a Zero inflated dynamic panel ordered probit model have been developed to address the above challenges. Average partial effects that presents the impacts on the specific probabilities per unit change in the covariates are also given. Since the solutions are not of closed form, Bayesian approach was used to estimate the parameters of the model since it allows sampling from the conditional distributions. Bayesian approach does not need a large sample to ensure the adequacy of asymptotic approximations and incorporate prior information in estimation. A Monte Carlo study was carried out to investigate some theoretical properties of the estimators in the models. The study found that the Zero inflated dynamic panel ordered probit models with independent and correlated error terms produced consistent estimators. The Zero inflated dynamic panel ordered probit models with independent and correlated error terms had more accurate estimators than the Dynamic panel ordered probit model. The Zero inflated dynamic panel ordered probit model with independent error terms fitted the National Longitudinal Survey of Youth 1997 better than Zero inflated dynamic panel ordered probit model with correlated error terms and Dynamic panel ordered probit model. The Zero inflated dynamic panel ordered probit model with correlated error terms fitted the National Longitudinal Survey of Youth 1997 better than Dynamic panel ordered probit model.

Keywords: Ordered probit (OP) model, zero inflated dynamic panel ordered probit (ZIDPOP) model, state dependence, unobserved heterogeneity and initial condition problem

1. Introduction

Some categorical datasets are characterized by zero inflation due to non-participation decision and zero consumption of substances such as drug. For instance, the first zero may be recorded from a respondent who has inelastic demand for cigarettes while the second zero may be recorded from a respondent who is currently not smoking due to low income but are likely to become smoker in case their income is raised. Overlooking the sources of these zeros may leads to model misspecification and concealing differential effects explanatory variables have on the two zeros as reported by Gurmu and Dagne (2009) ^[7]. Harris and Zhao (2007) ^[8] proposed the Zero Inflated Ordered Probit (ZIOP) model that permit the splitting of the observed non-users into non-participants and zero consumption potential users. The study revealed that the use of OP model would confuse the effects of some significant covariates that have contrasting impacts on the two zeros. The ZIOP Model allowed for the recognition of covariates that are significant in each source of zero and are key to policy making. Gurmu and Dagne (2009) ^[7] extended ZIOP model to Bayesian approach and applied the model in modelling tobacco consumption. Oh *et al.* (2012) ^[13] proposed a Bayesian analysis of sequential ZIOP model and applied it to alcohol consumption in Korea. Their analysis showed that about 30% of non-drinkers are genuine non-drinkers. A study on the marginal effect of each covariate showed that certain covariates have effects on the genuine non-drinkers and potential drinkers in converse directions, which may not be easily revealed by an OP model.

According to Yong-Woo (2016) ^[18], analysing the dynamics of individuals' health status involves the recognition of true state dependence in the persistence of the health status. The persistence arise from the tendency of the past health status being a significant determinant of the current health status. Therefore, the observed heterogeneity, unobserved heterogeneity and true state dependence should be cautiously structured in order to properly recognize how health in the subsequent period is accounted for by current health. According to Heckman (1981c) ^[9], the poor treatment of unobserved individual effects gives rise to a conditional association between past and current experiences, which is referred to as spurious state dependence. To separate these dependences is significant in policy-making. For instance, if true state dependence is present in health evolution, then short-term health interventions tend to improve individual health in the long-term. However, if unobserved individual effect is related over time and not carefully controlled for, then the previous health status may seem to be a determinant of current health, solely because it is a proxy for such temporally persistent unmeasured variables. In this case, short-term health interventions have no effect on the longer-term health status.

Contoyannis *et al.* (2004) ^[4] proposed the Dynamic panel ordered probit (DPOP) model based on Wooldridge (2005) ^[17] approach for initial observation and used it to study self-assessed health based on British Household Panel Survey for the 1991-1998 period. Wooldridge (2005) ^[17] suggested an unobserved individual effects conditional on the initial values and exogenous variables. Specifying the distribution of unobserved individual effects on these variables can lead to very tractable functional forms and consistent estimators for dynamic panel data with random effects. They found a substantial positive state dependence and unobserved individual effects. Including state dependence considerably lowers the impact of observed individual heterogeneity. Conditioning on the first period health outcomes and within-individual averages of the exogenous variables lowers the impact of heterogeneities and state dependence. In their model, unobservable individual effect accounts for around 30% of the unexplained variation in health. Yong-Woo (2016) ^[18] also proposed DPOP model based on Rabe-Hesketh and Skrondal (2013) ^[15] approach for initial observation and used it to study self-assessed health based on Korea Labour and Income Panel Study for the 2003-2012 period. Their results revealed that the health dynamics in the country are characterized by significant positive state dependence and 20% unobserved individual effect. The effect of many covariates vanishes when the state dependence and unobserved individual effects are controlled for. They also found a significant positive association between the initial period health observation and unobserved latent health. Therefore, this indicated that it is essential to control for self-assessed health at the beginning of observations.

The Semi-parametric Zero inflated dynamic panel probit model with selectivity proposed by Christelis and Galdeano (2009) ^[3] based on Akay (2009) ^[1] approach for initial conditions, non-parametric distribution for unobserved individual effects and known cut points may results in biased and inconsistent estimates. The constrained model proposed by Akay (2009) ^[1] and used by Christelis and Galdeano (2009) ^[3] to model the initial condition was shown by Rabe-Hesketh and Skrondal (2013) ^[15] to be severely biased. It implicitly sets the coefficients of the initial covariates equal to the coefficients for the subsequent periods, which is at odds with the form of the correct distribution. The main intention is that the conditional distribution of the unobserved effect, given the covariates at all periods (including the first observation), relies more directly on the initial-period explanatory variables than on the covariates at the other periods-in some cases it depends only on the initial-period covariates and the first response variable. The coefficients of the initial-period covariates should therefore not be constrained to be equal to the coefficients at the other periods. They showed that the bias for the constrained model practically vanishes when the initial period covariates are included as additional covariate or by using Wooldridge's original auxiliary model. The model cannot be applied in data with unknown cut points or without heaping. They ignored the association between the unobserved individual effects that is used to determine whether the factors affecting the unobserved individual effects in participation decision are the same as the one affecting the unobserved individual effects at consumption levels. The model could not allow estimation of the inter-unit correlation coefficients that is used to determine the latent error variance associated with the unobserved heterogeneity. They used simulated maximum likelihood approach proposed by Lee and Oguzoglu (2007) ^[11] and Kano (2008) ^[10] where the individual's effects are integrated out by computing the double integral by simulation. Mulkay (2015) ^[12] pointed that this procedure could be time-consuming and imprecise even though we use modern simulator like Geweke Hajivassiliou-Keane or Halton simulators, because we need to compute R cumulative density function with a large value of R in order to obtain sufficient precision in the log-likelihood function.

This paper proposed Zero inflated dynamic panel ordered probit models with independent error terms (ZIDPOPI) and Zero inflated dynamic panel ordered probit models with correlated error terms (ZIDPOPC) that incorporates the state dependence, unobserved heterogeneities based on parametric distribution with selectivity, initial conditions based on Rabe-Hesketh and Skrondal (2013) ^[15] approach and unknown cut points. Since the solutions are not of closed form, Bayesian approach was used to estimate the parameters of the models since it allows sampling from the conditional distributions.

2. Zero Inflation Dynamic Panel Ordered Probit Model

Suppose y_{it}^b represents binary specification for either non-participation or participation for i^{th} respondent at time t while y_{it}^o denote ordinal discrete response for i^{th} respondent at time t where $i = 1, 2, \dots, n$ and $t = 0, 1, 2, \dots, T$. The respondents are sampled independently from the underlying population. The time $t = 0$ denotes the first period. y_{it}^b is related to a latent variable y_{it}^{b*} via the mapping: $y_{it}^b = 0$ for $y_{it}^{b*} \leq 0$ and $y_{it}^b = 1$ for $y_{it}^{b*} > 0$ at time t . The latent variables y_{it}^{b*} denotes the inclination for participation (smoking, change, symptoms) and is given in vector form by,

$$y_{it}^{b*} = \phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + u_{it}^b \quad (1)$$

ϕ_1 , γ_1 and β_1 are vectors of unknown coefficients. $u_{it}^b = \kappa_{1i} + e_{it}^b$ is the composite error term. e_{it}^b is considered to be strictly exogenous. κ_{1i} is the time invariant individual-specific fixed effect. To capture the state dependence, y_{it-1}^b is a vector of indicators for the individual's status in the previous wave. The asymptotic properties is with respect to a fixed T while $N \rightarrow \infty$.

Conditional on $y_{it}^b = 1$, the consumption (number of cigarettes smoked, change, severity of symptoms) levels are denoted by a discrete variable y_{it}^o ($y_{it}^o = 0, 1, 2, \dots, K$) at time t that is generated by an Ordered probit (OP) model through a second underlying latent variable through thresholding. It can also be expressed in vector form as,

$$y_{it}^{o*} = \phi_2 y_{it-1}^o + \gamma_2' z_{it} + \beta_2' v_i + u_{it}^o \quad (2)$$

$u_{it}^o = \kappa_{2i} + e_{it}^o$ is the composite error term. e_{it}^o is assumed to be strictly exogenous. κ_{2i} is the time invariant individual-specific fixed effect. To capture state dependence, y_{it-1}^o is a vector of indicators for the respondent's status in the previous wave and the model can be interpreted as a first order Markov process. The asymptotic properties is with respect to a fixed T while $N \rightarrow \infty$.

The latent variable y_{it}^{o*} and the observed variable y_{it}^o are connected by

$$y_{it}^o = k \Leftrightarrow \tau_{k-1} < y_{it}^{o*} \leq \tau_k \quad k = 0, 1, 2, \dots, K \quad (3)$$

where τ_k are cut points of ordinal response. In order to ensure that the cumulative distribution function for y_{it}^o is appropriately defined, we expect that $\tau_{k-1} < \tau_k \forall k$. We specify that $-\infty < \tau_{-1} < \tau_0 < \tau_1 < \dots < \tau_{K-1} < \tau_K < \infty$ where $\tau_{-1} = -\infty, \tau_K = \infty$ and $\tau_0 = 0$ for identification purpose and to avoid the handling of boundary parameters.

Assuming that u_{it}^o is standard Gaussian random variable, the probabilities are given by

$$P(y_{it}^o) = \begin{cases} P(y_{it}^o = 0 | \phi_2 y_{it-1}^o, \gamma_2' z_{it}, \beta_2' v_i, \kappa_{2i}, y_{it}^b = 1) = \Phi(-\phi_2 y_{it-1}^o - \beta_2' v_i - \gamma_2' z_{it} - \kappa_{2i}) & k = 0 \\ P(y_{it}^o = k | \phi_2 y_{it-1}^o, \gamma_2' z_{it}, \beta_2' v_i, \kappa_{2i}, y_{it}^b = 1) = \Phi(\tau_k - \phi_2 y_{it-1}^o - \beta_2' v_i - \gamma_2' z_{it} - \kappa_{2i}) \\ \quad - \Phi(\tau_{k-1} - \phi_2 y_{it-1}^o - \beta_2' v_i - \gamma_2' z_{it} - \kappa_{2i}) & k = 1, 2, \dots, K-1 \\ P(y_{it}^o = K | \phi_2 y_{it-1}^o, \gamma_2' z_{it}, \beta_2' v_i, \kappa_{2i}, y_{it}^b = 1) = \Phi(\phi_2 y_{it-1}^o + \gamma_2' z_{it} + \beta_2' v_i + \kappa_{2i} - \tau_{K-1}) & k = K \end{cases} \quad (4)$$

While y_{it}^b and y_{it}^o are not individually observable in terms of the zeros, they are observed via the criterion $y_{it} = y_{it}^b y_{it}^o$. That is, to observe $y_{it} = 0$ outcome we expect either that $y_{it}^b = 0$ (the individual is a nonparticipant) or jointly that $y_{it}^o = 0$ and $y_{it}^b = 1$ (the individual is a participant but with zero consumption). To observe a positive $y_{it} = 1$, we require jointly that the individual is a participant $y_{it}^b = 1$ and that $y_{it}^o > 0$ (the individual is a participant and with non-zero consumption).

Assuming that u_{it}^b and u_{it}^o are standard Gaussian random variables, identically and independently distributed, then the full probabilities for y_{it} is given by

$$P(y_{it}) = \begin{cases} P(y_{it} = 0 | x, z) = P(y_{it}^b = 0 | x) + P(y_{it}^b = 1 | x) P(y_{it}^o = 0 | x, z, y_{it}^b = 1) \\ P(y_{it} = k | x, z) = P(y_{it}^b = 1 | x) P(y_{it}^o = k | x, z, y_{it}^b = 1) & k = 1, 2, \dots, K \end{cases} \quad (5)$$

On the other hand, u_{it}^b and u_{it}^o could be assumed to be correlated since they correspond to the same individual. Thus, y_{it} is given by,

$$y_{it} = \begin{cases} 0 & \text{if } (y_{it}^{b*} \leq 0) \text{ or } (y_{it}^{b*} > 0 \text{ and } y_{it}^{o*} \leq 0) \\ k & \text{if } (y_{it}^{b*} > 0 \text{ and } \tau_{k-1} < y_{it}^{o*} \leq \tau_k) \quad k = 1, 2, \dots, K-1 \\ K & \text{if } (y_{it}^{b*} > 0 \text{ and } y_{it}^{o*} > \tau_{K-1}) \end{cases} \quad (6)$$

The corresponding probabilities are given by

$$P(y_{it}) = \begin{cases} P(y_{it} = 0 | x, z) = \left[1 - \Phi(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + \kappa_{1i}) \right] + \\ \quad \left[\Phi(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + \kappa_{1i}, -\phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - \kappa_{2i}, -\rho_{e_{it}^b e_{it}^o}) \right] \\ P(y_{it} = k | x, z) = \left[\Phi(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + \kappa_{1i}, \tau_k - \phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - \kappa_{2i}, -\rho_{e_{it}^b e_{it}^o}) - \right. \\ \quad \left. \Phi(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + \kappa_{1i}, \tau_{k-1} - \phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - \kappa_{2i}, -\rho_{e_{it}^b e_{it}^o}) \right] \\ \quad k = 1, 2, \dots, K-1 \\ P(y_{it} = K | x, z) = \left[\Phi(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + \kappa_{1i}, \phi_2 y_{it-1}^o + \gamma_2' z_{it} + \beta_2' v_i + \kappa_{2i} - \tau_{K-1}, \rho_{e_{it}^b e_{it}^o}) \right] \end{cases} \quad (7)$$

where $\Phi(f, g; \rho)$ denotes the cumulative standardized bivariate normal distribution with correlation coefficient ρ between the two bivariate random elements.

Therefore, the chance of a zero observation has been “inflated” as it is a blend of the probability of “non-participation” from the split probit model and the probability of “zero consumption” from the OP process. This study assumed a balanced panel model where information about a respondents and required variables are reported at each wave.

Rabe-Hesketh and Skrondal (2013)^[15] specifies the individual time invariant error term κ_{1i} and κ_{2i} , as normally distributed terms and includes the initial observation y_{i0}^{b*} and y_{i0}^{o*} , initial observation of the covariates x_{i0} and z_{i0} while the time averages of the time varying covariates $\bar{x}_i = \frac{1}{T} \sum_{t=1}^T x_{it}$ and $\bar{z}_i = \frac{1}{T} \sum_{t=1}^T z_{it}$ such that:

$$\kappa_{1i} = h_1^b y_{i0}^{b*} + h_0^b + h_2^{b*} \bar{x}_i + h_3^{b*} x_{i0} + \delta_{1i} \quad (8)$$

$$\kappa_{2i} = h_1^o y_{i0}^{o*} + h_0^o + h_2^{o*} \bar{z}_i + h_3^{o*} z_{i0} + \delta_{2i} \quad (9)$$

The individual effects are considered, in every period, to be linear in the strictly exogenous covariates and the initial conditions. With

$$\kappa_{1i} | (y_{i0}^{b*}, \bar{x}_i, x_{i0}) \sim N(h_1^b y_{i0}^{b*} + h_0^b + h_2^{b*} \bar{x}_i + h_3^{b*} x_{i0}, \sigma_1^2)$$

$$\kappa_{2i} | (y_{i0}^{o*}, \bar{z}_i, z_{i0}) \sim N(h_1^o y_{i0}^{o*} + h_0^o + h_2^{o*} \bar{z}_i + h_3^{o*} z_{i0}, \sigma_2^2)$$

$$\delta_{1i} | (y_{i0}^{b*}, \bar{x}_i, x_{i0}) \sim N(0, \sigma_1^2) \text{ and } \delta_{2i} | (y_{i0}^{o*}, \bar{z}_i, z_{i0}) \sim N(0, \sigma_2^2). \quad h_0^b, h_1^b, h_2^b, h_3^b, h_0^o, h_1^o, h_2^o \text{ and } h_3^o \text{ are parameters to be}$$

estimated. δ_{1i} and δ_{2i} are uncorrelated with $(y_{i0}^b, \bar{x}_i, x_{i0})$ and $(y_{i0}^o, \bar{z}_i, z_{i0})$ respectively. The scalars h_1^b and h_1^o indicate the dependence of the individual effects on the initial conditions.

Substituting equation (8) into equation (1) leads to a final underlying latent variable specification:

$$y_{it}^{b*} = \phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_1^b y_{i0}^b + h_0^b + h_2^{b*} \bar{x}_i + h_3^{b*} x_{i0} + \delta_{1i} + e_{it}^b \quad (10)$$

$$\text{with } e_{it}^b | (y_{it-1}^b, x_{it}, y_{i0}^b, \bar{x}_i, w_i, x_{i0}, \delta_{1i}) \sim N(0, 1)$$

Substituting equation (9) into equation (2) leads to a final underlying latent variable specification:

$$y_{it}^{o*} = \phi_2 y_{it-1}^o + \gamma_2' z_{it} + \beta_2' v_i + h_1^o y_{i0}^o + h_0^o + h_2^{o*} \bar{z}_i + h_3^{o*} z_{i0} + \delta_{2i} + e_{it}^o \quad (11)$$

$$\text{with } e_{it}^o | (y_{it-1}^o, z_{it}, y_{i0}^o, \bar{z}_i, v_i, z_{i0}, \delta_{2i}) \sim N(0, 1)$$

The vectors (e_{it}^b, e_{it}^o) and $(\delta_{1i}, \delta_{2i})$ were considered to be uncorrelated with each other, and independently and identically distributed over time and across individuals. They follows a bivariate normal distribution with mean zero and covariance matrices below.

$$\Sigma_{\epsilon_{beo}} = \begin{pmatrix} \sigma_{eb}^2 & \rho_{ebeo} \sigma_{eb} \sigma_{eo} \\ \rho_{ebeo} \sigma_{eb} \sigma_{eo} & \sigma_{eo}^2 \end{pmatrix} = \begin{pmatrix} 1 & \rho_{ebeo} \\ \rho_{ebeo} & 1 \end{pmatrix} \quad \Sigma_{\delta_{1i}\delta_{2i}} = \begin{pmatrix} \sigma_1^2 & \rho_{\delta_{1i}\delta_{2i}} \sigma_1 \sigma_2 \\ \rho_{\delta_{1i}\delta_{2i}} \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}$$

Where $\sigma_{eb}^2 = \sigma_{eo}^2 = 1$ for identification purpose. The entries of the covariance matrices are also estimated. All time dummies are ignored from x_i and z_i to evade the perfect collinearity. The estimates of h_1^b and h_1^o are significant as they reveal the association between the individual effect and initial observation.

2.1 Bayesian Approach to Zero inflated dynamic ordered probit model

Conditioned on the unobserved individual effects κ , the responses y_{it} ($t = 1, 2, \dots, T$) are considered independent. Then, the contribution to the likelihood for individual i , conditional on the covariates and the unobserved individual effects, would be the joint probability given by

$$p(y_{it} | \text{others}) = \prod_{t=1}^T \prod_{k=0} \left[\begin{aligned} & \left(1 - \Phi \left(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_0^b + h_1^b y_{i0}^b + h_2^{b*} \bar{x}_i + h_3^{b*} x_{i0} + \delta_{1i} \right) \right) + \\ & \Phi \left(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_0^b + h_1^b y_{i0}^b + h_2^{b*} \bar{x}_i + h_3^{b*} x_{i0} + \delta_{1i}, \right. \\ & \quad \left. -\phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - h_0^o - h_1^o y_{i0}^o - h_2^{o*} \bar{z}_i - h_3^{o*} z_{i0} - \delta_{2i}, -\rho \right) \end{aligned} \right] \quad (12)$$

$$\prod_{k>0}^K \left[\begin{aligned} & \Phi \left(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_0^b + h_1^b y_{i0}^b + h_2^{b*} \bar{x}_i + h_3^{b*} x_{i0} + \delta_{1i}, \right. \\ & \quad \left. \tau_k - \phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - h_0^o - h_1^o y_{i0}^o - h_2^{o*} \bar{z}_i - h_3^{o*} z_{i0} - \delta_{2i}, -\rho \right) \\ & -\Phi \left(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_0^b + h_1^b y_{i0}^b + h_2^{b*} \bar{x}_i + h_3^{b*} x_{i0} + \delta_{1i}, \right. \\ & \quad \left. \tau_{k-1} - \phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - h_0^o - h_1^o y_{i0}^o - h_2^{o*} \bar{z}_i - h_3^{o*} z_{i0} - \delta_{2i}, -\rho \right) \end{aligned} \right]$$

where Φ is the standard normal distribution function.

Let $\begin{pmatrix} e_{it}^b \\ e_{it}^o \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_{e_{beo}}\right)$ and $\kappa = \begin{pmatrix} \kappa_{1i}^b \\ \kappa_{2i}^o \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_{\delta_i \delta_{2i}}\right)$. Let $\Theta = \{\phi_1, \phi_2, \gamma_1, \beta_1, \gamma_2, \beta_2, \Sigma_{iebeo}, \Sigma_{e_{beo}}, \Sigma_{\delta_{1i}\delta_{2i}}, \tau\}$ be a vector of unknown parameters. The prior distribution of the random effects κ is assumed to have a hierarchical structure. Then, the prior distributions are specified as follows:

$$p(\Theta) = p(\phi_1) p(\phi_2) p(\gamma_1) p(\beta_1) p(\gamma_2) p(\beta_2) p(\kappa_{1i} | 0, \Sigma_{\delta_i \delta_{2i}}) p(\kappa_{2i} | 0, \Sigma_{\delta_i \delta_{2i}}) p(\Sigma_{\delta_{1i}\delta_{2i}}) p(\tau) p(\Sigma_{e_{beo}}) \quad (13)$$

We have adopted a Bayesian hierarchical model approach for estimation and inference in the ZIDPOP models with exogenous variable selection. Therefore, we have to choose a prior for each parameter in the model. The prior specifications for most of the parameters are straightforward in the sense that we set them relatively non-informative and conjugate. The main motivation for employing the non-informative priors is that these priors would not dictate the inference of the parameters involved in the model. The Metropolis-Hastings algorithm was used due to the inability to sample from the full conditional distributions. The joint posterior distribution of the parameters conditional on the data was obtained by combining the likelihood function and the specified prior distributions via Bayes' theorem, as:

$$p(\Theta | y_{it}, x, z, w, v) \propto p(\Theta) \prod_{i=1}^n \prod_{t=1}^T \prod_{k=0}^K \left[\left(1 - \Phi\left(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_0^b + h_1^b y_{i0}^b + h_2^{tb} \bar{x}_i + h_3^{tb} x_{i0} + \delta_{1i}\right) \right) + \right. \\ \left. \Phi\left(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_0^b + h_1^b y_{i0}^b + h_2^{tb} \bar{x}_i + h_3^{tb} x_{i0} + \delta_{1i}, \right. \right. \\ \left. \left. -\phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - h_0^o - h_1^o y_{i0}^o - h_2^{to} \bar{z}_i - h_3^{to} z_{i0} - \delta_{2i}, -\rho \right) \right]^{d_{ik}} \\ \prod_{i=1}^n \prod_{t=1}^T \prod_{k>0}^K \left[\Phi\left(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_0^b + h_1^b y_{i0}^b + h_2^{tb} \bar{x}_i + h_3^{tb} x_{i0} + \delta_{1i}, \right. \right. \\ \left. \left. \tau_k - \phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - h_0^o - h_1^o y_{i0}^o - h_2^{to} \bar{z}_i - h_3^{to} z_{i0} - \delta_{2i}, -\rho \right) \right]^{d_{ik}} \\ \left. - \Phi\left(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_0^b + h_1^b y_{i0}^b + h_2^{tb} \bar{x}_i + h_3^{tb} x_{i0} + \delta_{1i}, \right. \right. \\ \left. \left. \tau_{k-1} - \phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - h_0^o - h_1^o y_{i0}^o - h_2^{to} \bar{z}_i - h_3^{to} z_{i0} - \delta_{2i}, -\rho \right) \right] \quad (14)$$

Where d_{itk} is an indicator function such that $d_{itk} = 1$ if $d_{itk} = k$ and 0 otherwise.

Let $\theta_t = \phi_1, \phi_2, \gamma_1, \gamma_2, \beta_1$ and β_2 . We assume non-informative normal priors for regression coefficients θ_t , with mean θ_t^* and variance Ω_{θ_t} , that is, $\theta_t \sim N(\theta_t^*, \Omega_{\theta_t})$ which are selected to make the distribution proper, that is, integrates to 1 but diffuse with large variances. The prior distributions are given by,

$$p(\theta_t) \propto \Omega_{\theta_t}^{-1/2} \exp\left\{-\frac{1}{2}(\theta_t - \theta_t^*)' \Omega_{\theta_t}^{-1} (\theta_t - \theta_t^*)\right\} \quad (15)$$

Then the full conditional posterior distribution required to implement the MCMC algorithm is given by,

$$p(\theta | x, z, w, v, \Theta_{-\theta}) \propto |\Omega_{\theta}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\theta - \theta^*)' \Omega_{\theta}^{-1} (\theta - \theta^*)\right) \times \\ \prod_{i=1}^n \prod_{t=1}^T \prod_{k=0}^K \left[\left(1 - \Phi\left(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_0^b + h_1^b y_{i0}^b + h_2^{tb} \bar{x}_i + h_3^{tb} x_{i0} + \delta_{1i}\right) \right) + \right. \\ \left. \Phi\left(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_0^b + h_1^b y_{i0}^b + h_2^{tb} \bar{x}_i + h_3^{tb} x_{i0} + \delta_{1i}, \right. \right. \\ \left. \left. -\phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - h_0^o - h_1^o y_{i0}^o - h_2^{to} \bar{z}_i - h_3^{to} z_{i0} - \delta_{2i}, -\rho \right) \right]^{d_{ik}} \\ \prod_{i=1}^n \prod_{t=1}^T \prod_{k>0}^K \left[\Phi\left(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_0^b + h_1^b y_{i0}^b + h_2^{tb} \bar{x}_i + h_3^{tb} x_{i0} + \delta_{1i}, \right. \right. \\ \left. \left. \tau_k - \phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - h_0^o - h_1^o y_{i0}^o - h_2^{to} \bar{z}_i - h_3^{to} z_{i0} - \delta_{2i}, -\rho \right) \right]^{d_{ik}} \\ \left. - \Phi\left(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_0^b + h_1^b y_{i0}^b + h_2^{tb} \bar{x}_i + h_3^{tb} x_{i0} + \delta_{1i}, \right. \right. \\ \left. \left. \tau_{k-1} - \phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - h_0^o - h_1^o y_{i0}^o - h_2^{to} \bar{z}_i - h_3^{to} z_{i0} - \delta_{2i}, -\rho \right) \right] \quad (16)$$

The proposal distribution for each of the variables of interest given above was normal distribution that is symmetrical. The probability of acceptance is given by,

$$\alpha(x^{(i)} | x^{(i-1)}) = \min\left(1, \frac{\pi(x^{(i)})}{\pi(x^{(i-1)})}\right) \quad (17)$$

Since $q(x^{(i)} | x^{(i-1)}) = q(x^{(i-1)} | x^{(i)})$ for symmetric distribution. Let $u \sim \text{uniform}(0, 1)$. If $u < \alpha$ then accept the proposal: $x^{(i)} \leftarrow x^{cand}$ else reject the proposal: $x^{(i)} \leftarrow x^{(i-1)}$.

The prior for variance –covariance matrix $\Sigma_{\delta_{1i}\delta_{2i}} = \begin{pmatrix} \sigma_1^2 & \rho_{\delta_{1i}\delta_{2i}} \sigma_1 \sigma_2 \\ \rho_{\delta_{1i}\delta_{2i}} \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}$ is inverse Wishart distributions denoted by

$\Sigma \sim IW(v, \Lambda)$ with density.

$$p(\Sigma) \propto |\Sigma|^{-\frac{v+d+1}{2}} \exp\left(-\frac{1}{2} \text{tr}(\Lambda \Sigma^{-1})\right) \quad (18)$$

Λ was a positive definite d dimensional matrix and v is a scalar value representing degrees of freedom. In order to get a proper prior we should set $v > d - 1$. At a multivariate level the prior mean is $E(\Sigma) = \frac{\Lambda}{v - d - 1}$. Using an inverse Wishart prior for the

covariance matrix induces an inverse scale chi-square distribution for each variance $\sigma_i \sim \text{inv}\chi^2\left(v - d + 1, \frac{\lambda_{ii}}{v - d + 1}\right)$ where λ_{ii} it's a diagonal entry of Λ . The conditional distribution of the correlation coefficient is given by

$$p(\rho | \sigma_1^2, \sigma_2^2) \propto (1 - \rho^2)^{-\frac{v+d+1}{2}} \exp\left(-\frac{\lambda_{22}\sigma_1^2 + \lambda_{11}\sigma_2^2 - 2\lambda_{12}\sigma_1\sigma_2}{\sigma_1^2\sigma_2^2(1 - \rho^2)}\right). \quad (19)$$

Assuming an $IW(v_0, D_0)$ where $v_0 = 2$ and $D_0 = I$, I is an identity matrix, draw $\Sigma_{\delta_{1i}\delta_{2i}}$ from its full conditional

$$\pi(\Sigma_{\delta_{1i}\delta_{2i}} | \text{others}) = \pi(\Sigma_{\delta_{1i}\delta_{2i}} | \kappa_i) = IW(n + v_0, D_0 + \kappa_i' \kappa_i) \quad (20)$$

We assume non-informative normal priors for individual random effect. The prior for the individual random effect in binary and

ordinal part was $\kappa_i = \begin{pmatrix} \kappa_{1i} \\ \kappa_{2i} \end{pmatrix} \sim N_2(0, \Sigma_{\delta_{1i}\delta_{2i}})$ where $\Sigma_{\delta_{1i}\delta_{2i}} = \begin{pmatrix} \sigma_1^2 & \rho_{\delta_{1i}\delta_{2i}} \sigma_1 \sigma_2 \\ \rho_{\delta_{1i}\delta_{2i}} \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}$ and $\rho_{\delta_{1i}\delta_{2i}}$ represent the correlation

between κ_{1i} and κ_{2i} . To specify the Metropolis algorithm we specify the proposal distribution $h_k(\kappa)$, from which potential new values are drawn. If we choose f_k that is in the exponential family as the proposal distribution then the acceptance function takes a particularly neat form. Let κ denote the previous draw from the conditional distribution of $\kappa | y$ and generate a new value, κ^* , for the k^{th} component of κ using the proposal distribution. If we denote $\kappa^* = (\kappa_1, \kappa_2, \dots, \kappa_{i-1}, \kappa_i^*, \kappa_{i+1}, \dots, \kappa_i)$, then we accept κ^* as the new value with probability $\alpha_i(\kappa, \kappa^*)$ and otherwise we retain κ . Here $\alpha_i(\kappa, \kappa^*)$ is given by

$$\alpha_i(\kappa, \kappa^*) = \min\left(1, \frac{f_{\kappa|y}(\kappa^* | y, \text{others}) h_k(\kappa)}{f_{\kappa|y}(\kappa | y, \text{others}) h_k(\kappa^*)}\right) \quad (21)$$

Upon choosing $h_k = f_k$, the second term in braces in the above equation simplifies to

$$\frac{f_{\kappa|y}(\kappa^* | y, \text{others}) h_k(\kappa)}{f_{\kappa|y}(\kappa | y, \text{others}) h_k(\kappa^*)} = \frac{\prod_{i=1}^n \prod_{t=1}^T f_{y|k}(y_{it} | \kappa^*, \text{others}) f_k(\kappa^* | \Sigma_{\delta_{1i}\delta_{2i}}) f_k(\kappa | \Sigma_{\delta_{1i}\delta_{2i}})}{\prod_{i=1}^n \prod_{t=1}^T f_{y|k}(y_{it} | \kappa, \text{others}) f_k(\kappa | \Sigma_{\delta_{1i}\delta_{2i}}) f_k(\kappa^* | \Sigma_{\delta_{1i}\delta_{2i}})} = \frac{\prod_{i=1}^n \prod_{t=1}^T f_{y|k}(y_{it} | \kappa^*, \text{others})}{\prod_{i=1}^n \prod_{t=1}^T f_{y|k}(y_{it} | \kappa, \text{others})} \quad (22)$$

The only unknown parameter in Θ is ρ , the correlation between e_{it}^b and e_{it}^o . The values lies within $-1, 1$. Therefore, the choice for prior distribution for ρ can be uniform $(-1, 1)$ or a proper distribution based on reparametrization. Let v denote the hyperbolic arc-tangent transformation of ρ , that is, $v = \text{atanh}(\rho)$ and taking hyperbolic tangent transformation of v returns us back to $\rho = \tanh(v)$. We assume that $v \sim N(v^*, \sigma_v^2)$.

Full conditional posterior distribution required to implement the MCMC algorithm is given by,

$$f(v | \text{others}) \propto \sigma_v^{-1} \exp\left\{-\frac{(v - v^*)^2}{2\sigma_v^2}\right\} \prod_{i=1}^n \prod_{t=1}^T \prod_{k=0}^K \left[\left(1 - \Phi\left(\phi_{1i} y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_0^b + h_1^b y_{i0}^b + h_2^{tb} \bar{x}_i + h_3^{tb} x_{i0} + \delta_{1i}\right)\right) + \right. \\ \left. \Phi\left(\phi_{1i} y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_0^b + h_1^b y_{i0}^b + h_2^{tb} \bar{x}_i + h_3^{tb} x_{i0} + \delta_{1i}, \right. \right. \\ \left. \left. -\phi_{2i} y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - h_0^o - h_1^o y_{i0}^o - h_2^{to} \bar{z}_i - h_3^{to} z_{i0} - \delta_{2i}, -\rho\right) \right]^{d_{ik}} \\ \prod_{i=1}^n \prod_{t=1}^T \prod_{k=0}^K \left[\Phi\left(\phi_{1i} y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_0^b + h_1^b y_{i0}^b + h_2^{tb} \bar{x}_i + h_3^{tb} x_{i0} + \delta_{1i}, \right. \right. \\ \left. \left. \tau_k - \phi_{2i} y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - h_0^o - h_1^o y_{i0}^o - h_2^{to} \bar{z}_i - h_3^{to} z_{i0} - \delta_{2i}, -\rho\right) \right. \\ \left. -\Phi\left(\phi_{1i} y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_0^b + h_1^b y_{i0}^b + h_2^{tb} \bar{x}_i + h_3^{tb} x_{i0} + \delta_{1i}, \right. \right. \\ \left. \left. \tau_{k-1} - \phi_{2i} y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - h_0^o - h_1^o y_{i0}^o - h_2^{to} \bar{z}_i - h_3^{to} z_{i0} - \delta_{2i}, -\rho\right) \right]^{d_{ik}} \quad (23)$$

Finally, in choosing prior distributions for the threshold parameters, τ 's, caution is needed because of the order restriction on them. Sampling from the complete conditional distribution for τ 's results in slow mixing. Instead, we use Cowles' (1996)^[5] algorithm to simulate τ 's. Let $i=1$ and $\sigma_{MH} = 0.05/K$. This value of σ_{MH} provides a starting point, and adjustments to σ_{MH} may be vital if suitable acceptance rates for τ are not attained. Generate a candidate g for updating τ^{k-1} : For $j=1, \dots, K-1$, sample $g_j \sim N(\tau_j^{i-1}, \sigma_{MH}^2)$ truncated to the interval $(g_{j-1}, \tau_{j+1}^{i-1})$ (take $g_{-1} = -\infty$, $g_0 = 0$ and $g_K = \infty$). Calculate the acceptance ratio R given by

$$R = \prod_{i=1}^n \prod_{t=1}^T \frac{\Phi\left(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_0^b + h_1^b y_{i0}^b + h_2^b \bar{x}_i + h_3^b x_{i0} + \delta_{1i}, g_{y_i} - \phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - h_0^o - h_1^o y_{i0}^o - h_2^o \bar{z}_i - h_3^o z_{i0} - \delta_{2i}\right)}{\Phi\left(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_0^b + h_1^b y_{i0}^b + h_2^b \bar{x}_i + h_3^b x_{i0} + \delta_{1i}, \tau_{y_i} - \phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - h_0^o - h_1^o y_{i0}^o - h_2^o \bar{z}_i - h_3^o z_{i0} - \delta_{2i}\right)} \times \frac{\prod_{j=1}^{K-1} \Phi\left((\tau_{j+1}^{i-1} - \tau_j^{i-1}) / \sigma_{MH}\right) - \Phi\left((g_{j-1} - \tau_j^{i-1}) / \sigma_{MH}\right)}{\prod_{j=1}^{K-1} \Phi\left((g_{j+1} - g_j) / \sigma_{MH}\right) - \Phi\left((\tau_{j-1}^{i-1} - g_j) / \sigma_{MH}\right)} \quad (24)$$

Set $\tau^{(i)} = g$ with probability R . Otherwise, take $\tau^{(i)} = \tau^{(i-1)}$. The first term denotes the contribution from the likelihood function. The second term accounts for the difference in the normalization of the proposal densities on the truncated normal intervals from which candidate points are drawn. When implementing this algorithm, the acceptance rate for the cut points should be monitored. If this rate falls too much below 25% or above 50%, σ_{MH} should be lowered or raised, respectively. The parameter were estimated using Single Component Metropolis-Hastings algorithm.

2.2 Model Selection

The model selection procedure was performed using Deviance Information Criterion (DIC) proposed by Spiegelhalter et al. (2002)^[15]. It is given by $DIC = \text{goodness} - \text{of} - \text{fit} + \text{penalty for complexity}$

$$DIC = \bar{D}(\theta) + p_D \quad (25)$$

where $\bar{D}(\theta) = E(D(\theta) | y)$ is the posterior mean of the deviance, $D(\theta)$, and $p_D = \bar{D}(\theta) - \hat{D}(\theta) = E[D(\theta) | y] - D[E(\theta | y)]$ is the difference in the posterior mean of the deviance and the deviance evaluated at the posterior mean of the parameters. The model with a smaller DIC should be preferred to models with larger DIC. Models are penalized both by the value of $\bar{D}$, which favours a better fit, but also by the effective number of parameters p_D . The advantage of DIC compared to other criteria such as Akaike Information Criterion is that the DIC is simply obtained from the MCMC samples. DIC is estimated from posterior samples:

$$DIC = 2\bar{D} - D(\bar{\theta}) \quad (26)$$

Where $\bar{\theta} = \frac{1}{L} \sum_{t=1}^L \theta^{(t)}$ and $\bar{D} = \frac{1}{L} \sum_{t=1}^L -2 \log p(y | \theta^{(t)})$. L in the number of the iterations.

2.3 Average Partial Effects

Average partial effects (APEs) are used to indicate the size of associations between ordinal observation and the covariates. The average partial effects are computed by scaling the coefficient vector. The partial effects are averaged over the population distribution of heterogeneity and computed using the population averaged parameters γ_a . In the random effects specifications

these are given by $\gamma_a = \frac{\gamma}{\sqrt{1 + \sigma^2}}$. The sign of the APE has a clear qualitative interpretation, with a positive sign inferring a positive

association with ordered response and vice versa. In the zero inflated dynamic ordered probit model it is possible to compute APEs for each of the two categories of participation and categories of consumption. We obtain the average partial effects of covariate $\omega = (x_i, x_{it}, z_i, z_{it})$ on various probabilities assuming the regression errors follow the bivariate normal distribution.

If ω is a binary regressors, the marginal effect of ω on probability, say P , is the difference in the probability evaluated at 1 and 0, conditional on observable values of covariates: $P(\omega_i = 1) - P(\omega_i = 0)$. For continuous explanatory variables, the marginal effect is given by the partial derivative of the probability of interest with respect to ω_i . Wooldridge (2002)^[17] shows that computing the partial effect at the observed values of the regressors for each observation and averaging the estimates over the observations provides a consistent estimate of the APEs.

The average partial effects on the probability of participation was given by;

$$APE_i(\lambda_{\omega_i}, y_{it}^{b*} = 1) = \frac{\lambda_{\omega_i}}{NT} \sum_{l=1}^N \sum_{t=1}^T \phi(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_0^b + h_1^b y_{i0}^b + h_2^b \bar{x}_i) \quad (27)$$

where $\varphi(\cdot)$ is the probability density function of the standard normal distribution and λ_{ω_i} was the coefficient in the inflation part associated with variable ω_i .

Let $u_{1,qh} = \phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_1^b y_{i0}^b + h_0^b + h_2^b \bar{x}_i + h_3^b x_{i0}$ and $W = \phi_2 y_{it-1}^o + \gamma_2' z_{it} + \beta_2' v_i + h_1^o y_{i0}^o + h_0^o + h_2^o \bar{z}_i + h_3^o z_{i0}$.

In terms of the zeros category, the effect on the probability of non-participation (zero inflation) was

$$APE_i(\lambda_{\omega_i}, y_{it}^{b*} = 0) = \frac{\lambda_{\omega_i}}{NT} \sum_{l=1}^N \sum_{t=1}^T -\varphi(u_{1,qh}) \quad (28)$$

while

$$APE_i(\lambda_{\omega_i}, \beta_{\omega_i}, y_{it}^{b*} = 1, y_{it}^{o*} = 0) = \frac{1}{NT} \sum_{l=1}^N \sum_{t=1}^T \left\{ \Phi\left(\frac{-u_{1,qh0} + \rho_{ebeo} u_{1,qh}}{\sqrt{1 - \rho_{ebeo}^2}}\right) \varphi(u_{1,qh}) \lambda_{\omega_i} - \left[\Phi\left(\frac{u_{1,qh} - \rho_{ebeo} u_{1,qh0}}{\sqrt{1 - \rho_{ebeo}^2}}\right) \varphi(-u_{1,qh0}) \beta_{\omega_i} \right] \right\} \quad (29)$$

represents the marginal effect on the probability of zero consumption.

The average partial effects on the probability of observing zero inflation was obtained as a sum of the marginal effects in (28) and (29); that is,

$$APE_i(y_{it}^{o*} = 0) = APE_i(y_{it}^{b*} = 0) + APE_i(y_{it}^{b*} = 1, y_{it}^{o*} = 0) \quad (30)$$

Let $u_{1,qhk} = \tau_k - W$, $u_{1,qhk-1} = \tau_{k-1} - W$ and $u_{1,qhk} = W^* - \tau_K$. The average partial effects for the remaining choices are:

$$APE_i(\lambda_{\omega_i}, \beta_{\omega_i}, y_{it}^{b*} = 1, y_{it}^{o*} = 1) = \sum_{i=1}^N \sum_{t=1}^T \left[\Phi\left(\frac{(\tau_2 - W) + \rho_{ebeo} u_{1,qh}}{\sqrt{1 - \rho_{ebeo}^2}}\right) - \Phi\left(\frac{-W + \rho_{ebeo} u_{1,qh}}{\sqrt{1 - \rho_{ebeo}^2}}\right) \right] \varphi(u_{1,qh}) \lambda_{\omega_i} \\ - \left[\Phi\left(\frac{u_{1,qh} + \rho_{ebeo} (\tau_2 - W)}{\sqrt{1 - \rho_{ebeo}^2}}\right) \varphi(\tau_2 - W) - \Phi\left(\frac{u_{1,qh} - \rho_{ebeo} W}{\sqrt{1 - \rho_{ebeo}^2}}\right) \varphi(-W) \right] \beta_{\omega_i} \quad (31)$$

$$APE_i(\lambda_{\omega_i}, \beta_{\omega_i}, y_{it}^{b*} = 1, y_{it}^{o*} = 2) = \sum_{i=1}^N \sum_{t=1}^T \left[\Phi\left(\frac{(\tau_3 - W) + \rho_{ebeo} u_{1,qh}}{\sqrt{1 - \rho_{ebeo}^2}}\right) - \Phi\left(\frac{\tau_2 - W + \rho_{ebeo} u_{1,qh}}{\sqrt{1 - \rho_{ebeo}^2}}\right) \right] \varphi(u_{1,qh}) \lambda_{\omega_i} \\ - \left[\Phi\left(\frac{u_{1,qh} + \rho_{ebeo} (\tau_3 - W)}{\sqrt{1 - \rho_{ebeo}^2}}\right) \varphi(\tau_3 - W) - \Phi\left(\frac{u_{1,qh} + \rho_{ebeo} (\tau_2 - W)}{\sqrt{1 - \rho_{ebeo}^2}}\right) \varphi(\tau_2 - W) \right] \beta_{\omega_i} \quad (32)$$

$$APE_i(\lambda_{\omega_i}, \beta_{\omega_i}, y_{it}^{b*} = 1, y_{it}^{o*} = k) = \sum_{i=1}^N \sum_{t=1}^T \left[\Phi\left(\frac{(\tau_k - W) + \rho_{ebeo} u_{1,qh}}{\sqrt{1 - \rho_{ebeo}^2}}\right) - \Phi\left(\frac{\tau_{k-1} - W + \rho_{ebeo} u_{1,qh}}{\sqrt{1 - \rho_{ebeo}^2}}\right) \right] \varphi(u_{1,qh}) \lambda_{\omega_i} \\ - \left[\Phi\left(\frac{u_{1,qh} + \rho_{ebeo} (\tau_k - W)}{\sqrt{1 - \rho_{ebeo}^2}}\right) \varphi(\tau_k - W) - \Phi\left(\frac{u_{1,qh} + \rho_{ebeo} (\tau_{k-1} - W)}{\sqrt{1 - \rho_{ebeo}^2}}\right) \varphi(\tau_{k-1} - W) \right] \beta_{\omega_i} \quad (33)$$

$$APE_i(\lambda_{\omega_i}, \beta_{\omega_i}, y_{it}^{b*} = 1, y_{it}^{o*} = K) = \sum_{i=1}^N \sum_{t=1}^T \left[\Phi\left(1 - \frac{(\tau_K - W) + \rho_{ebeo} u_{1,qh}}{\sqrt{1 - \rho_{ebeo}^2}}\right) \varphi(u_{1,qh}) \lambda_{\omega_i} + \left[\Phi\left(\frac{u_{1,qh} + \rho_{ebeo} (\tau_K - W)}{\sqrt{1 - \rho_{ebeo}^2}}\right) \varphi(\tau_K - W) \right] \beta_{\omega_i} \right] \quad (34).$$

The asymptotic standard errors of the $APE_{k,i,t}$ are computed using the Delta method as the square roots of the diagonal elements of

$$Var(APE) = \nabla_{\Theta} APE_{k,i,t} Var(\Theta) \nabla_{\Theta} APE'_{k,i,t} \quad (35)$$

3. Results and Discussions

3.1 Data Simulation

This section describe the data generating process that was used to simulate the data. In order to generate the initial values problem, we operated the scheme for twenty-five periods before data were observed. The number of individuals was ($n = 750$), the initial

period ($s = -25$) and the number of periods for the current simulation, so that each individual was observed from -25 to 36 . We generated the right hand side variables by drawing values for the covariates, the individual fixed effects and the idiosyncratic errors for both binary and ordinal models. The initial outcome was assumed to follow a Bernoulli trial with probability 0.5 for Dynamic panel binary probit model. The initial outcome was assumed to follow a binomial trial with probability 0.5 for Dynamic panel ordered probit model.

The following parameters were used for simulation. The parameters for binary model were fixed at $\phi_1 = 0.5$, $\gamma_{11} = -1.0$, $\beta_{11} = -1.0$, $\sigma_1 = 1.0$. The inter-period correlation was $\sigma_1^2 / (\sigma_1^2 + 1) = 0.5$. The parameters for ordinal model were fixed at $\phi_2 = 0.5$, $\gamma_{21} = 0.6$, $\gamma_{22} = 1.5$, $\beta_{21} = 1.0$. The inter-period correlation was $\sigma_2^2 / (\sigma_2^2 + 1) = 0.5$. The covariance matrix was given by $\Sigma_{\delta_{it}\delta_{2it}} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$ and the correlation matrix was given by $\Sigma_{\varepsilon b \varepsilon o} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$. The number of replication (R) was 100.

Models performance were evaluated using the Root Mean Square Error (RMSE). Let θ be the true parameter. Let $\hat{\theta}_k^r$ be the estimates of k^{th} parameter from r^{th} Monte Carlo replication, R be the total number of Monte Carlo replications, and θ_k be the true population value of the k^{th} parameter from the model. The Monte Carlo mean of estimates of the estimated k^{th} parameter was calculated as follows $\bar{\theta}_k = \frac{1}{R} \sum_{r=1}^R \hat{\theta}_k^r$. To evaluate the accuracy of the estimates, which capture both bias and variability, Root Mean

Square Error (RMSE) were calculated for each parameter. The RMSE was given by $\bar{\theta}_k = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\theta}_k^r - \theta)^2}$. RMSE is always positive. A value of zero, though never achieved in practice, would indicate a perfect fit. In practice, a parameter with a lower value of RMSE is more accurate than one with a higher value.

3.2 Results for Bayesian Inference in Simulated Data

The MCMC simulation was run for 40,000 iterations. The first 10,000 samples were discarded as the burn-in period. The thinning interval was six. The acceptance rates were maintained between 20% and 30% except for the cut points. The acceptance rates were maintained between 25% and 50% for the cut points as proposed by Cowles (1996) [5].

Table 1: The results based on $n=750$ and $T=10$ for DPOP, ZIDPOPI and ZIDPOPC models

Models		DPOP		ZIDPOPI		ZIDPOPC	
Par	True	Est	SD	Est	SD	Est	SD
h_0^b				-0.638	0.088	-0.712	0.100
ϕ_1	0.5			0.496	0.060	0.517	0.062
γ_{11}	-1.0			-1.005	0.039	-1.062	0.040
β_1	-1.0			-0.732	0.093	-0.754	0.100
h_1^b				1.278	0.104	1.329	0.110
h_{2m}^b				0.048	0.168	0.031	0.166
h_{2i}^b				0.291	0.055	0.299	0.054
σ_1	1.0			0.922	0.032	0.938	0.033
h_0^o		-1.181	0.070	-0.586	0.098	-0.742	0.088
ϕ_2	0.5	0.196	0.025	0.498	0.042	0.498	0.040
γ_{21}	0.4	0.171	0.017	0.434	0.029	0.403	0.028
γ_{22}	1.0	0.403	0.018	1.078	0.037	1.055	0.036
β_2	0.7	-0.534	0.083	0.610	0.099	0.420	0.101
h_1^o		0.652	0.050	1.030	0.071	0.973	0.069
h_{2m1}^o		-0.001	0.135	0.052	0.168	-0.028	0.166
h_{2m2}^o		-0.035	0.141	0.176	0.176	0.170	0.162
h_{2i1}^o		-0.077	0.044	-0.177	0.052	-0.133	0.051
h_{2i1}^o		-0.252	0.045	-0.414	0.058	-0.393	0.054

σ_2	1.0	0.860	0.029	0.904	0.032	0.924	0.033
$\rho_{e_{beo}}$	0.5					0.431	0.067
$\rho_{\delta_{1i}\delta_{2i}}$	0.5			0.413	0.041	0.435	0.040
τ_1	2.6	1.156	0.026	2.816	0.083	2.731	0.082
τ_2	4.2	2.141	0.043	4.514	0.109	4.429	0.109
DIC	17146.5		8658.465		8633.924		

The Table 1 indicate the parameters, true values, estimates, standard errors and DIC values. There is a moderate correlation between the error terms in ZIDPOPC model of 0.431 implying that the factors affecting the participation decision are the same as the one affecting the consumption levels. There is a moderate correlation between the unobserved individual effects in ZIDPOPC model of 0.435 implying that the factors affecting the unobserved individual effects in participation decision are the same as the one affecting the unobserved individual effects at consumption levels. The variance of the individual effects in participation decision was 0.938. This indicated that 46.80% of the latent error variance is associated with unobserved individual effects, as evaluated by the intra unit correlation coefficient in participation decision. The variance of the individual effects for the decision on consumption levels was 0.924. This indicated that 46.06% of the latent error variance is attributable to unobserved individual effects, as evaluated by the intra-unit correlation coefficient at consumption levels. The Deviance Information Criteria (DIC) was given as a measure of goodness-of-fit. The model with the smallest DIC was considered to provide a better fit than the rest. The ZIDPOPC model clearly fitted the better than ZIDPOPI and DPOP models. The ZIDPOPI model clearly fitted the better than the DPOP model.

3.3 Assessing the Accuracy of the Estimators in the Models

The accuracy of the three models was compared using RMSE.

Table 2: Comparison of DPOP, ZIDPOPI and ZIDPOPC when $n=750$ and $T=10$ based on RMSE from the simulated data.

Models		DPOP		ZIDPOPI		ZIDPOPC	
Par	True	Ests	RMSE	Ests	RMSE	Ests	RMSE
ϕ_1	0.5			0.496	0.035	0.517	0.057
γ_{11}	-1.0			-1.005	0.036	-1.062	0.073
β_1	-1.0			-0.732	0.307	-0.754	0.073
σ_1	1.0			0.922	0.078	0.938	0.063
ϕ_2	0.5	0.196	0.305	0.498	0.059	0.498	0.042
γ_{21}	0.4	0.171	0.206	0.434	0.053	0.403	0.020
γ_{22}	1.0	0.403	0.805	1.078	0.085	1.055	0.065
β_2	0.7	-0.534	0.298	0.610	0.187	0.420	0.303
σ_2	1.0	0.860	0.140	0.904	0.096	0.924	0.078
$\rho_{e_{beo}}$	0.5					0.431	0.073
$\rho_{\delta_{1i}\delta_{2i}}$	0.5			0.413	0.088	0.435	0.067
τ_1	2.6	1.156	1.444	2.816	0.223	2.731	0.148
τ_2	4.2	2.141	2.059	4.514	0.336	4.429	0.254

Although the ZIDPOPI and ZIDPOPC models have more parameters than DPOP model, for the comparison purpose, comments were restricted to shared parameters, that is, ϕ_2 , γ_{21} , γ_{22} , β_2 , σ_2 , τ_1 and τ_2 . The majority of MSE for the ZIDPOPI and ZIDPOPC models were lower than the DPOP model. The individual RMSE of all of the parameters (6 out of 7 compared), were lower in the ZIDPOPI and ZIDPOPC models than the DPOP model. This indicated that the ZIDPOPI and ZIDPOPC estimates were more accurate than DPOP estimates.

3.4 Assessing the Consistency of the Estimators in the Models

The Table 3 and Table 4 gives the results from the models when $n=350$ and $T=10$ and $n=750$ and $T=10$ respectively.

Table 3: The results based on $n=350$ and $T=10$ for DPOP, ZIDPOPI and ZIDPOPC models

Models		DPOP		ZIDPOPI		ZIDPOPC	
Par	True	Ests	SD	Ests	SD	Ests	SD
ϕ_1	0.5			0.482	0.091	0.532	0.089

γ_{11}	-1.0			-1.053	0.059	-1.056	0.058
β_1	-1.0			-0.757	0.144	-0.779	0.145
σ_1	1.0			0.933	0.048	0.943	0.049
ϕ_2	0.5	0.201	0.034	0.511	0.061	0.540	0.060
γ_{21}	0.4	0.172	0.025	0.406	0.041	0.404	0.041
γ_{22}	1.0	0.419	0.026	1.061	0.056	1.078	0.055
β_2	0.7	-0.490	0.124	0.594	0.152	0.447	0.150
σ_2	1.0	0.861	0.043	0.916	0.047	0.926	0.048
$\rho_{e_{beo}}$	0.5					0.430	0.095
$\rho_{\delta_{1i}\delta_{2i}}$	0.5			0.415	0.061	0.433	0.058
τ_1	2.6	1.130	0.037	2.718	0.125	2.787	0.129
ϕ_1	4.2	2.120	0.062	4.453	0.169	4.511	0.175

Table 4: The results based on n=750 and T=10 for DPOP, ZIDPOPI and ZIDPOPC models

Models		DPOP		ZIDPOPI		ZIDPOPC	
Par	True	Ests	SD	Ests	SD	Ests	SD
ϕ_1	0.5			0.496	0.060	0.517	0.062
γ_{11}	-1.0			-1.005	0.039	-1.062	0.040
β_1	-1.0			-0.732	0.093	-0.754	0.100
σ_1	1.0			0.922	0.032	0.938	0.033
ϕ_2	0.5	0.196	0.025	0.498	0.042	0.498	0.040
γ_{21}	0.4	0.171	0.017	0.434	0.029	0.403	0.028
γ_{22}	1.0	0.403	0.018	1.078	0.037	1.055	0.036
β_2	0.7	-0.534	0.083	0.610	0.099	0.420	0.101
σ_2	1.0	0.860	0.029	0.904	0.032	0.924	0.033
$\rho_{e_{beo}}$	0.5					0.431	0.067
$\rho_{\delta_{1i}\delta_{2i}}$	0.5			0.413	0.041	0.435	0.040
τ_1	2.6	1.156	0.026	2.816	0.083	2.731	0.082
ϕ_1	4.2	2.141	0.043	4.514	0.109	4.429	0.109

The study compares estimates in Table 3 when n=350 and T=10 and Table 4 when n=750 and T=10. The results from these two tables indicated that as n increases from 350 to 750, the estimates tend to the true values. For instance, the true value of ϕ_1 was 0.5, when n=350, ϕ_1 was 0.532 for ZIDPOPC model and when n=750, ϕ_1 was 0.517. This showed that as n was increased, the estimates tend to the true value. This indicated that ZIDPOPC model's estimators were consistent.

3.5 Application of the Models to Real Life Data

The main dataset used in this thesis was the National Longitudinal Survey of Youth 1997 (NLSY97). We eliminated the respondents with missing data for both responses and covariates. We constructed smoking intensities from three questions from the NLSY97. The zero consumption was created from two questions: "Have you ever smoked a cigarette?" and "During the last 30 days, how many days did you smoke a cigarette?" individuals who continuously answered "No" to the first question have perfectly inelastic demand for cigarettes, and for these individuals zero demands for cigarettes are optimal choices. Individuals who answer "None" to the second question belong to one of these two categories: those whose optimal choice of zero consumption are defined by corner solutions and infrequent smokers. A positive consumption was created from the question "When you smoked during the last 30 days, how many cigarettes did you usually smoke each day?" Let y_{it} be a measures of tobacco consumptions, by an individual i at time t . y_{it} is number of sticks of cigarettes smoked on days that individual i smokes, including zero consumption. For example, in the drinking sample, $i = 1, 2, \dots, 2500$ and $t = 0, 1, 2, \dots, 8$. Describing the frequency distribution of the crude y_{it} measurements does not actually aid in condensing the smoking intensities for empirical analysis as some respondents reported to have smoked 99 sticks of cigarette on the days they smoked. Transforming the crude

number of cigarettes smoked to 0-3 ordinal one-unit interval is popular in smoking literature. Thus, the study constructed four ordinal outcomes (0-3) of smoking intensities from their respective observed quantities y_{it} .

This can be summarized as follows;

$$y_{it} = \begin{cases} 0 & \text{if individual } i \text{ is not a current smoker or has never smoked} \\ 1 & \text{if individual } i \text{ smokes weekly or less} \\ 2 & \text{if individual } i \text{ smokes daily and smokes } < 20 \text{ sticks} \\ 3 & \text{if individual } i \text{ smokes daily and smokes } \geq 20 \text{ sticks} \end{cases}$$

The ordinal values $y_{it} = 0, 1, 2, 3$, stands for zero, low, moderate and high levels of smoking intensities.

3.6 Results for Bayesian Inference in Smoking Data

The MCMC simulation was run for 21,000 iterations. The first 3,000 samples were discarded as the burn-in period. The thinning interval was equal to four. The convergences were analyzed using trace and density plots. The acceptance rates were maintained between 20% and 30%. The acceptance rates were maintained between 25% and 50% for the cut points as proposed by Cowles (1996) ^[5].

Table 5: The results for n=2500 and T=8 for DPOP, ZIDPOPI and ZIDPOPC models in Smoking Data.

MODELS	DPOP		ZIDPOPI		ZIDPOPC	
Param	Ests	SD	Ests	SD	Ests	SD
ϕ_1			1.417	0.061	1.475	0.074
Age			0.002	0.037	-0.005	0.037
Edu			0.090	0.047	0.092	0.050
Gender			0.083	0.070	0.099	0.071
Race			-0.313	0.098	-0.306	0.095
h_1^b			1.533	0.083	1.533	0.076
σ_1			0.986	0.019	0.987	0.019
ϕ_2	0.547	0.016	0.399	0.020	0.411	0.021
Age	0.071	0.015	0.085	0.024	0.089	0.023
Edu	0.032	0.014	0.012	0.016	0.013	0.016
Gender	0.186	0.050	0.262	0.071	0.257	0.066
h_1^o	0.932	0.034	0.708	0.039	0.706	0.039
σ_2	1.023	0.019	0.976	0.019	0.975	0.018
$\rho_{e_{beo}}$					0.116	0.080
$\rho_{\delta_i \delta_{2i}}$			0.468	0.021	0.470	0.022
τ_1	1.347	0.020	1.958	0.054	1.958	0.051
τ_2	2.455	0.028	3.195	0.063	3.192	0.059
DIC	39812.15		23040.9		23075.09	

Table 5 shows estimates standard errors and DIC values for the three models. The variance of individual effects for participation was 0.986. This indicated that 49.30% of the latent error variance is associated to individual effect, as evaluated by the intra-unit correlation coefficient in smoking decision. The variance of individual effects for consumption levels was 0.976. This indicated that 48.79% of the latent error variance is associated to unobserved individual effect, as evaluated by the intra-unit correlation coefficient at consumption levels. There was a weak correlation coefficient of 0.116 between the error terms. This implied that the variables affecting the participation are different from the one affecting consumption levels. There was a moderate correlation coefficient of 0.468 between the individual effects. This implied that the variables affecting the individual effects at participation were the same as the one affecting individual effects at consumption levels.

The DIC was given as a measure of goodness-of-fit. The model with the smallest DIC was considered to fit the data better than the rest. The ZIDPOPI model clearly fitted the data better than the ZIDPOPC and DPOP model. This is attributed to the very weak correlation between the error terms in ZIDPOPC model. The ZIDPOPC model clearly fitted the data better than the DPOP model.

Table 6: Average Partial Effects for variables in Correlated Zero Inflated Dynamic Panel Ordered Probit Model (ZIDPOPC) for Real data

Param		Ests	Std error	t_value	P_value
ϕ_1	$y_{it}^b = 0$	0.353	0.026	13.465	0.000
	$y_{it}^b = 1$	0.342	0.025	13.465	0.000
Age	$y_{it}^b = 0$	0.001	0.000	26.737	0.000
	$y_{it}^b = 1$	-0.001	0.000	-26.737	0.000
Edu	$y_{it}^b = 0$	-0.012	0.001	-20.079	0.000

	$y_{it}^b = 1$	0.013	0.001	20.079	0.000
Gender	$y_{it}^b = 0$	0.019	0.001	14.014	0.000
	$y_{it}^b = 1$	0.019	0.001	14.014	0.000
Race	$y_{it}^b = 0$	-0.055	0.005	-10.503	0.000
	$y_{it}^b = 1$	-0.061	0.006	-10.503	0.000
ϕ_2	$y_{it}^0 = 0$	-0.031	0.001	-47.156	0.000
	$y_{it}^0 = 1$	0.031	0.001	47.156	0.000
	$y_{it}^0 = 2$	0.039	0.001	47.156	0.000
	$y_{it}^0 = 3$	0.016	0.000	47.156	0.000
Age	$y_{it}^0 = 0$	-0.007	0.000	-42.983	0.000
	$y_{it}^0 = 1$	0.000	0.000	-42.983	0.000
	$y_{it}^0 = 2$	0.011	0.000	42.983	0.000
	$y_{it}^0 = 3$	0.017	0.000	42.983	0.000
Edu	$y_{it}^0 = 0$	-0.006	0.000	-62.240	0.000
	$y_{it}^0 = 1$	0.011	0.000	62.240	0.000
	$y_{it}^0 = 2$	0.004	0.000	62.240	0.000
	$y_{it}^0 = 3$	0.004	0.000	62.240	0.000
Gender	$y_{it}^0 = 0$	0.008	0.001	11.055	0.000
	$y_{it}^0 = 1$	0.419	0.028	15.139	0.000
	$y_{it}^0 = 2$	0.196	0.013	15.139	0.000
	$y_{it}^0 = 3$	0.022	0.001	15.139	0.000
ϕ_1	$y_{it}^0 = 0$	-0.031	0.006	-5.489	0.000
Age	$y_{it}^0 = 0$	-0.007	0.000	-47.156	0.000
Edu	$y_{it}^0 = 0$	0.006	0.000	42.983	0.000
Gender	$y_{it}^0 = 0$	-0.011	0.000	-62.240	0.000

Table 6 shows the estimated average partial effects, t values, p values and 95% confidence intervals. All the covariates are statistically significant at 1%. The covariates with positive sign implying a positive association with the participation and consumption at all levels and vice versa. The average partial effects for the state dependence in nonparticipation decision and participation decision were 0.353 and 0.342 respectively. This indicated that as one move from the previous year to the next, the probability of non-smoking and smoking would increase by 35.3% and 34.2% respectively. The average partial effects for the age in non-participation decision and participation decision were 0.001 and -0.001 respectively. This indicated that single-unit rise in age will provide a 0.1% increase in the chance of non-smoking for an otherwise “average” individual while one-unit increase in age will produce a -0.1% decrease in the probability of smoking for an otherwise “average” individual. The average partial effects for the state dependence, at various consumption levels, that is, zero, low, moderate and high consumption levels were -0.055, 0.031, 0.039 and 0.016, respectively. This indicated that as one move from the previous year to the next, the probability of various consumption levels, that is, zero, low, moderate and high would change by -5.5%, 3.1%, 3.9% and 1.6% respectively. The average partial effects for the age, at various consumption levels, that is, zero, low, moderate and high consumption levels were -0.007, 0.000, 0.011 and 0.017, respectively. This indicated that one-unit increase in age will produce -0.7%, 0.0%, 1.1% and 1.7% change in various consumption levels, that is, zero, low, moderate and high consumption levels respectively. The average partial effects for the state dependence, age, education and gender, at various zero consumption level were -0.031, -0.007, 0.006 and -0.011, respectively. This indicated that as one change from female to male, the probability of various consumption levels, that is, zero, low, moderate and high would change by -3.1%, -0.7%, 0.6% and -1.1% respectively.

4. Conclusion

The aim of this study was to develop ZIDPOP model with both correlated and uncorrelated error terms and compared them with DPOP model. The Deviance information criteria from the simulation studies showed that ZIDPOPC model provided a better than ZIDPOPI and ZIDPOPC models while ZIDPOPI model provided a better than DPOP model. ZIDPOPI and ZIDPOPC models had a smaller RMSE compared to DPOP model. This showed that ZIDPOPC and ZIDPOPI models' estimates were more accurate than DPOP model's estimates. The estimates of the ZIDPOPI and ZIDPOPC models tended to the true parameter values as $n \rightarrow \infty$. This indicated that ZIDPOPI and ZIDPOPC models estimators were consistent. The study also concluded that the state dependence should not be ignored since both participation decision and consumption levels were characterized by substantial positive state dependence. The presence of state dependence means that short-term policy interventions designed to reduce participation and consumption levels may have longer-term implications. The study found positive associations between the initial period participation decision and consumption levels observations and unobserved latent participation decision and consumption levels. Therefore, this indicated that it is crucial to control for participation decision and consumption levels at the initial period.

5. References

1. Akay A. Finite-sample comparison of alternative methods for estimating dynamic panel data models. *Journal of Applied Econometrics*. 2012;27(7):1189-204.
2. Çağlayan-Akay E, Van MH. Determinants of the Levels of Development Based on the Human Development Index: Bayesian Ordered Probit Model. *International Journal of Economics and Financial Issues*. 2017;7(5):425-431.

3. Christelis D, Galdeano SA. Smoking persistence across countries: An analysis using semi-parametric dynamic panel data models with selectivity. *Journal of Applied Econometrics*. 2009;15:334-360.
4. Contoyannis P, Jones A, Rice R. The dynamics of health in the British household panel study. *Journal of Applied Econometrics*. 2004;19(4):473-503.
5. Cowles MK. Accelerating Monte Carlo Markov chain convergence for cumulative- link generalized linear models. *Statistics and Computing*. 1996;6(2):10-111.
6. Gurmu S, Dagne AG. Bayesian Approach to Zero-Inflated Bivariate Ordered Probit Regression Model, with an application to tobacco use. *Journal of Probability and Statistics*. 2012;5:1-26.
7. Gurmu S, Dagne GA. Bayesian Approach to Zero-Inflated Ordered Probit Models, with an Application. In *Seminar on Bayesian Inference in Econometrics and Statistics*. St. Louis: Washington University in St. Luis, 2009.
8. Harris MN, Zhao X. A Zero-inflated Ordered Probit Model, with an Application to Modelling Tobacco Consumption. *Journal of Econometrics*. 2007;141(2):1073-1099.
9. Heckman J. Heterogeneity and state dependence. In *Studies in Labor Markets*, Rosen (ed). University of Chicago Press; c1981, p. 91-140.
10. Kano S. Like husband, like wife: A bivariate dynamic probit analysis of spousal obesities. Osaka Prefecture University, 2008.
11. Lee WS, Oguzoglu U. Well-being and ill- being: A bivariate panel data analysis, IZA Discussion Paper Series No. 3108. 2007, 30.
12. Mulkay B. Bivariate probit estimation for panel data: a two-step Gauss Hermite quadrature approach with an application to product and process innovations for France. University's de Montpellier, Montpellier, 2015.
13. Oh M, Oh HT, Park S. Bayesian Analysis of Korean Alcohol Consumption Data Using a Zero-Inflated Ordered Probit Model. *Korean Journal of Applied Statistics*. 2012;25(2):363-376.
14. Spiegelhalter DJ, Best N, Carline B, Van DL. Bayesian method of model complexity and fit. *Journal of the Royal Statistical Society*. 2002;64:583-639.
15. Rabe-Hesketh S, Skrondal A. Avoiding biased versions of Wooldridge's simple solution to the initial conditions problem. *Economics Letters*. 2013;120(2):346-349.
16. Wooldridge JM. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*. 2005;20(1):39-54.
17. Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA. 2002;108(2):245-54.
18. Yong-Woo L. State dependence, unobserved heterogeneity, and health dynamics in Korea. *Hitotsubashi Journal of Economics*. 2016;57(2):195-221.