

Classical Approach to Zero-Inflated Dynamic Panel Ordered Probit Model with an Application in Drug Abuse

John Kung'u^{*}, Leo Odongo, Ananda Kube

Department of Mathematics and Actuarial Science, Kenyatta University, Nairobi, Kenya

Email address:

johnkungu08@yahoo.com (J. Kung'u), odongo.leo@ku.ac.ke (L. Odongo), kube.ananda@ku.ac.ke (A. Kube)

^{*}Corresponding author

To cite this article:

John Kung'u, Leo Odongo, Ananda Kube. Classical Approach to Zero-Inflated Dynamic Panel Ordered Probit Model with an Application in Drug Abuse. *American Journal of Theoretical and Applied Statistics*. Vol. 11, No. 2, 2022, pp. 58-74. doi: 10.11648/j.ajtas.20221102.11

Received: January 20, 2022; **Accepted:** March 4, 2022; **Published:** March 12, 2022

Abstract: The Zero inflated ordered categorical data with time series structure are often a characteristic of behavioral 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 models 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, Maximum likelihood estimation based on Newton Raphson algorithm was used to estimate the parameters of the model. 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 correlated error terms fitted the National Longitudinal Survey of Youth 1997 better than Zero inflated dynamic panel ordered probit model with independent error terms and 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 the Dynamic panel ordered probit model.

Keywords: Correlated Zero Inflated Dynamic Panel Ordered Probit Model (ZIDPOPC), State Dependence, Unobserved Heterogeneity and Initial Condition Problem

1. Introduction

Some categorical datasets are characterized by zero inflation due to non-participation and zero consumption. 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 two types of zeros for non-users or users with zero consumption may leads to model misspecification and concealing differential effects covariates have on the two regimes, non-users versus users at all levels of consumption as reported by Gurmu and Dagne

[1]. Harris and Zhao [2] proposed the Zero inflated ordered probit (ZIOP) models that allow for the split of the observed non-users into non-participants and zero consumption potential users. They found that the use of a conventional ordered probit model would confuse the effects of some important explanatory variables that have opposing impacts on the two schemes. The proposed model allowed for the identification of variables that are important in each regime. This is potentially very important for policy analysis. Several researches have been conducted using the ZIOP model. They include Downward et al. [3] in modelling sports participation, Oh et al. [4] modelling Korean alcohol consumption, Bagozzi et al. [5] in Modelling Two Types of Peace in international Conflict Research, Yuan et al. [6] in modeling mushroom

consumption in the USA and Jinzhong et al. [7] in analyzing Hazmat truck drivers' violation behavior and associated risk factors.

According to Yong-Woo [8], analyzing the dynamics of individuals' health status involves the recognition of true state dependence in the persistence of the health status. The persistence arises 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 [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 [10] proposed the Dynamic panel ordered probit model based on Akay [11] approach for initial observation and used it to study self-assessed health based on British Household Panel Survey for the 1991-1998 period. Wooldridge [12] 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 permanent heterogeneity. The presence of state dependence means that short-term policy interventions designed to improve health may have longer-term implications. Including state dependence dramatically reduces the impact of observed individual heterogeneity. Conditioning on the initial period health outcomes and within-individual averages of the exogenous variables reduces the impact of heterogeneity and state dependence. In their model, unobservable heterogeneity accounts for around 30% of the unexplained variation in health. Yong-Woo [8] also proposed Dynamic panel ordered probit model based on Rabe-Hesketh and Skrondal [13] approach for initial observation and used it to study self-assessed health based on Korea Labor and Income Panel Study for the 2003-2012 period. Their results showed that health dynamics in Korea are characterized by significant positive state dependence and 20% unobserved heterogeneity. The explanatory power of many socioeconomic variables disappears if state dependence and unobserved heterogeneity 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 necessary to control for self-assessed health at the beginning of observations.

Raymond et al. [14] proposed a method to implement maximum likelihood estimation of the Dynamic panel data Type 2 and 3 Tobit models. The method entails writing the likelihood function conditional on the individual effects that are then integrated out with respect to their joint normal distribution. The likelihood function involves a two dimensional indefinite integral evaluated using "two-step" Gauss-Hermite quadrature. A Monte Carlo study shows that the quadrature works well in finite sample for a number of evaluation points as small as two. Their results showed that ignoring the unobserved heterogeneity, or the dependence between the initial conditions and the unobserved heterogeneity results in an overestimation of the coefficients of the state dependence. The two-step Gauss-Hermite quadrature approach was also used by Mulkay [15] to study the Bivariate panel binary probit model with an application to product and process innovations in France. The Gauss-Hermite Quadrature procedure was 40% faster than the simulated maximum likelihood procedures performed on the same dataset and on the same computer. A simulation showed the importance of estimating the correlation in random effects and the correlation between both equations. These correlations were used to assess the effects of unobserved heterogeneity on each equation.

The Semi-parametric Zero inflated dynamic panel probit model with selectivity proposed by Christelis and Galdeano [16] based on Akay [11] approach for initial conditions, non-parametric distribution for unobserved individual effects and known cut points may result in biased and inconsistent estimates. The constrained model proposed by Akay [11] and used by Christelis and Galdeano [16] to model the initial condition was shown by Rabe-Hesketh and Skrondal [13] to be severely biased because it implicitly sets the coefficients of the initial explanatory variables equal to the coefficients for the subsequent periods, which is at odds with the form of the correct distribution. The reason is that the conditional distribution of the unobserved effect, given the explanatory variables at all periods (including the initial period), depends more directly on the initial-period explanatory variables than on the explanatory variables at the other periods—in some cases it depends only on the initial-period explanatory variables and the initial dependent variable. The coefficients of the initial-period explanatory variables 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 disappears when the initial period explanatory variables are included as additional regressors or by using Wooldridge's original auxiliary model. The model cannot be applied in data with unknown cut points or without heaping. The model ignored the correlation coefficient 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 inter-unit

correlation coefficients that is used to determine the latent error variance attributable to unobserved heterogeneity. They used Simulated maximum likelihood approach proposed by Lee and Oguzoglu [17] and Kano [18] where the individual's effects are integrated out by computing the double integral by simulation. Mulkay [15] 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 that incorporates the state dependence, unobserved heterogeneities based on parametric distribution with selectivity, initial conditions based on Rabe-Hesketh and Skrondal [13] approach and unknown cut points. This paper used an alternative approach by Raymond et al. [14] based on a two-step Gauss-Hermite Quadrature to compute the double integral. Since the solutions are not of closed form, parameters of the models were estimated by maximum likelihood estimation based on Newton Raphson algorithm.

2. Zero Inflation Dynamic Panel Ordered Probit Model

Let y_{it}^b represents binary specification for either non-participation or participation for respondent i at time t while y_{it}^o denote ordinal discrete response of respondent i at time t where $i=1, 2, \dots, n$ and $t=0, 1, 2, \dots, T$. The respondents are sampled independently from the target population. The time $t=0$ denotes an initial 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*} and y_{i0}^{b*} denotes the propensity for participation (smoking, change, symptoms) and are given in vector form by,

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

$$y_{i0}^{b*} = \gamma_{01}' x_{i0} + \beta_{01}' w_i + u_{i0}^b \quad (2)$$

ϕ_1 , γ_{01} , γ_1 , β_{01} and β_1 are vectors of unknown coefficients. $u_{it}^b = \kappa_{1i} + e_{it}^b$ is the composite error term. x_{it} and w_i are vectors of time variant covariates and time invariant covariates respectively considered to be strictly exogenous variables. e_{it}^b is considered to be strictly exogenous, that is, the x_{it} are independent of e_{it}^b for all t and s . The error terms of the model are assumed as $e_{it}^b \sim \text{iid } N(0, \sigma_{eb}^2)$ and $\sigma_{eb}^2 = 1$ for identification purpose.

κ_{1i} is the time invariant individual-specific fixed effect (also known as unobserved heterogeneity) affecting the decision to participation or not and is uncorrelated with covariates and is also assumed to be orthogonal to exogenous variables following the standard random-effects assumption. The correlation between two sequential error terms is $\text{corr}(u_{it}^b, u_{is}^b) = \sigma_{\kappa_{1i}}^2 / (\sigma_{\kappa_{1i}}^2 + 1)$ ($t, t=1, 2, \dots, T; t \neq s$). To capture state dependence, y_{it-1}^b 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 probability of participation $P(y_{it}^b = 1)$ and $P(y_{i0}^b = 1)$ are given by,

$$P(y_{it}^b = 1 | \phi_1 y_{it-1}^b, \gamma_1' x_{it}, \beta_1' w_i, \kappa_{1i}) = \Phi(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + \kappa_{1i}) \quad (3)$$

$$P(y_{i0}^b = 1 | \gamma_{01}' x_{i0}, \beta_{01}' w_i, \kappa_{1i}) = \Phi(\gamma_{01}' x_{i0} + \beta_{01}' w_i + \kappa_{1i}) \quad (4)$$

and, by symmetry, for non-participation

$$P(y_{it}^b = 0 | \phi_1 y_{it-1}^b, \gamma_1' x_{it}, \kappa_{1i}) = 1 - \Phi(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + \kappa_{1i}) \quad (5)$$

$$P(y_{i0}^b = 0 | \gamma_{01}' x_{i0}, \beta_{01}' w_i, \kappa_{1i}) = 1 - \Phi(\gamma_{01}' x_{i0} + \beta_{01}' w_i + \kappa_{1i}) \quad (6)$$

where $\Phi(\cdot)$ represents the standard normal cumulative distribution function.

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

$$y_{it}^{o*} = \phi_2 y_{it-1}^o + \gamma_2' z_{it} + \beta_2' v_i + u_{it}^o \quad (7)$$

$$y_{i0}^{o*} = \gamma_{02}' z_{i0} + \beta_{02}' v_i + u_{i0}^o \quad (8)$$

$u_{it}^o = \kappa_{2i} + e_{it}^o$ is the composite error term. z_{it} and v_i are vectors of time variant covariates and time invariant covariates respectively assumed to be strictly exogenous variables. e_{it}^o is assumed to be strictly exogenous, that is, the x_{it} are independent

of e_{is}^b for all t and s . The error terms of the model are assumed as $e_{it}^o \sim \text{iid } N(0, \sigma_{eo}^2)$ and $\sigma_{eo}^2 = 1$ for identification purpose. κ_{2i} is the time invariant individual-specific fixed effect. The inter class correlation between two sequential error terms is $\text{corr}(u_{it}^o, u_{is}^o) = \sigma_{\kappa_{2i}}^2 / (\sigma_{\kappa_{2i}}^2 + 1)$ ($t, t = 1, 2, \dots, T; t \neq s$). 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

$$y_{it}^o = k \Leftrightarrow \tau_{k-1} < y_{it}^{o*} \leq \tau_k \quad (9)$$

where $k = 0, 1, 2, \dots, K$.

where τ_k are cut points of ordinal response. In order to ensure that the cumulative distribution function for y_{it}^o is properly defined, we require 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 u_{it}^o is standard Gaussian random variable, the probabilities are given by

$$P(y_{it}^o) = \begin{cases} P(y_{it}^o = 0 | \phi_2 y_{it-1}^o, \gamma_2' z_{it}, \beta_2' v_i, \kappa_{2i}, y_{it}^b = 1) = \\ \Phi(-\phi_2 y_{it-1}^o - \beta_2' v_i - \gamma_2' z_{it} - \kappa_{2i}) \quad k = 0 \\ P(y_{it}^o = k | \phi_2 y_{it-1}^o, \gamma_2' z_{it}, \beta_2' v_i, \kappa_{2i}, y_{it}^b = 1) = \\ \Phi(\tau_k - \phi_2 y_{it-1}^o - \beta_2' v_i - \gamma_2' z_{it} - \kappa_{2i}) \\ - \Phi(\tau_{k-1} - \phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - \kappa_{2i}) \quad k = 1, 2, \dots, K-1 \end{cases} \quad (10)$$

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 require 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 identically and independently distributed and follows the standard Gaussian random variables, the full probabilities for y_{it} is given by

$$P(y_{it}) = \begin{cases} P(y_{it} = 0 | x, z) = P(y_{it}^b = 0 | x) \\ + P(y_{it}^b = 1 | x) P(y_{it}^o = 0 | x, z, y_{it}^b = 1) \\ P(y_{it} = k | x, z) = \\ P(y_{it}^b = 1 | x) P(y_{it}^o = k | x, z, y_{it}^b = 1) \\ k = 1, 2, \dots, K \end{cases} \quad (11)$$

In this way, the chance of a zero observation has been “inflated” as it is a combination of the probability of “zero consumption” from the OP process plus the probability of “non-participation” from the split probit model.

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,

$$y_{it} = \begin{cases} 0 & \text{if } (y_{it}^{b*} \leq 0) \text{ or } (y_{it}^{b*} > 0 \text{ and } y_{it}^{o*} \leq 0) \\ k & \text{if } (y_{it}^{b*} > 0 \text{ and } \tau_{k-1} < y_{it}^{o*} \leq \tau_k) \quad k = 1, 2, \dots, K-1 \\ K & \text{if } (y_{it}^{b*} > 0 \text{ and } y_{it}^{o*} > \tau_{K-1}) \end{cases} \quad (12)$$

while the corresponding probabilities are given by,

$$P(y_{it}) = \begin{cases} P(y_{it} = 0 | x, z) = \left[1 - \Phi(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + \kappa_{1i}) \right] + \\ \Phi\left(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + \kappa_{1i}, \right. \\ \left. -\phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - \kappa_{2i}, -\rho_{ebeo}\right) \\ P(y_{it} = k | x, z) = \left[\Phi\left(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + \kappa_{1i}, \right. \right. \\ \left. \left. \tau_k - \phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - \kappa_{2i}, -\rho_{ebeo}\right) - \right. \\ \left. \Phi\left(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + \kappa_{1i}, \right. \right. \\ \left. \left. \tau_{k-1} - \phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2' v_i - \kappa_{2i}, -\rho_{ebeo}\right) \right] \\ k = 1, 2, \dots, K-1 \\ P(y_{it} = K | x, z) = \left[\Phi\left(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + \kappa_{1i}, \right. \right. \\ \left. \left. \phi_2 y_{it-1}^o + \gamma_2' z_{it} + \beta_2' v_i + \kappa_{2i} - \tau_K, \rho_{ebeo}\right) \right] \end{cases} \quad (13)$$

where $\Phi(f, g; \rho)$ denotes the cumulative distribution function of the standardized bivariate normal distribution with correlation coefficient ρ between the two bivariate random elements. This paper assumes a balanced panel model where information about a respondents and required variables are reported at each wave.

The initial conditions problem is a methodological challenge that stems from random-effects approach. For the first observation of the panel (initial conditions), due to the fact that we do not have data for the previous state on y_{i1}^b and y_{i1}^o (no values for y_{i0}^b and y_{i0}^o) we are unable to evaluate $P(y_{i1}^b, y_{i1}^o | y_{i0}^{b*}, y_{i0}^{o*}, x_{i0}, z_{i0})$. By overlooking it in the individual likelihood, researchers also overlook the data generation process for the first observation of the panel. This implies that they consider the data generating process of the first observation of the panel to be exogenous or to be in equilibrium. These assumptions hold only if the individual random effects are degenerative. If this assumption is not met, the initial conditions are explained by the individual random effects and overlooking them results into inconsistent estimates. Wooldridge [12] addressed the issue by specifying a distribution for the individual effects conditional on the initial conditions and the strictly exogenous covariates. The likelihood function obtained in the Wooldridge approach has a similar structure for both dynamic and static versions of the nonlinear model. Rabe-Hesketh and Skrondal [13] 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} and 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:

$$\kappa_{1i} = h_1^b y_{i0}^{b*} + h_0^b + h_2^{b*} \bar{x}_i + h_3^{b*} x_{i0} + \delta_{1i} \quad (14)$$

$$\kappa_{2i} = h_1^o y_{i0}^{o*} + h_0^o + h_2^{o*} \bar{z}_i + h_3^{o*} z_{i0} + \delta_{2i} \quad (15)$$

The individual effects are assumed, in each period, to be linear in the strictly exogenous explanatory variables and the initial conditions. With

$$\kappa_{1i} | (y_{i0}^{b*}, \bar{x}_i, x_{i0}) \sim N(h_1^b y_{i0}^{b*} + h_0^b + h_2^{b*} \bar{x}_i + h_3^{b*} x_{i0}, \sigma_1^2) \quad (16)$$

$$y_{it}^{b*} = \phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_1^b y_{i0}^b + h_0^b + h_2^{b*} \bar{x}_i + h_3^{b*} x_{i0} + \delta_{1i} + e_{it}^b \quad (18)$$

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 (16) into equation (7) leads to a final underlying latent variable specification:

$$y_{it}^{o*} = \phi_2 y_{it-1}^o + \gamma_2' z_{it} + \beta_2' v_i + h_1^o y_{i0}^o + h_0^o + h_2^{o*} \bar{z}_i + h_3^{o*} z_{i0} + \delta_{2i} + e_{it}^o \quad (19)$$

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

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

$$\Sigma_{e_{it}} = \begin{pmatrix} \sigma_{eb}^2 & \rho_{e_{it}} \sigma_{eb} \sigma_{eo} \\ \rho_{e_{it}} \sigma_{eb} \sigma_{eo} & \sigma_{eo}^2 \end{pmatrix} = \begin{pmatrix} 1 & \rho_{e_{it}} \\ \rho_{e_{it}} & 1 \end{pmatrix}$$

Since $\sigma_{eb}^2 = \sigma_{eo}^2 = 1$ for identification purpose.

$$\Sigma_{\delta_{it}} = \begin{pmatrix} \sigma_1^2 & \rho_{\delta_{it}} \sigma_1 \sigma_2 \\ \rho_{\delta_{it}} \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}$$

The entries of the covariance matrices are also estimated. All time dummies are ignored in x_i and z_i to evade perfect collinearity. The estimates of h_1^b and h_1^o are key informative of the relationship between the individual effect and initial observation.

Independence is a key assumption in the derivation of an ordinary probit model, that is, the joint probability for the data equals the product of the marginal probabilities and log-likelihood is the sum of the specific log-likelihood contributions. However, this does not apply to serially

$$\kappa_{2i} | (y_{i0}^{o*}, \bar{z}_i, z_{i0}) \sim N(h_1^o y_{i0}^{o*} + h_0^o + h_2^{o*} \bar{z}_i + h_3^{o*} z_{i0}, \sigma_2^2) \quad (17)$$

where

$$\delta_{1i} | (y_{i0}^{b*}, \bar{x}_i, x_{i0}) \sim N(0, \sigma_1^2)$$

$$\delta_{2i} | (y_{i0}^{o*}, \bar{z}_i, z_{i0}) \sim N(0, \sigma_2^2)$$

Denotes $h_0^b, h_1^b, h_2^b, h_3^b, h_0^o, h_1^o, h_2^o$ and h_3^o parameters to be estimated. δ_{1i} and δ_{2i} are independent of $(y_{i0}^b, \bar{x}_i, x_{i0})$ and $(y_{i0}^o, \bar{z}_i, z_{i0})$ respectively.

The coefficients h_1^b and h_1^o depict the dependence of the unobserved individual effects on the initial observation.

Substituting equation (15) into equation (1) leads to a final underlying latent variable specification:

dependent data since y_{it} is a function of y_{it-1} , $P(y_{it} = k)$ is no longer independent of $P(y_{it-r} = k)$ in turn, the joint probability of the data is no longer the product of the time-specific probabilities, and the log-likelihood is no longer the sum of the time-specific log-likelihood contributions. As [9] points out, the assumption that $y_{it} = (y_{it}^b, y_{it}^o)$ depends on its once lagged value y_{it-1} but not on any of its other lags implies that the joint distribution of $y_{i1}, y_{i2}, \dots, y_{iT}$ conditional on y_{i0} , $\kappa_i = (\kappa_{1i}, \kappa_{2i})$, (x_{it}, z_{it}) and (w_i, v_i) can be written as

$$\prod_{i=1}^N \prod_{t=1}^T f(y_{it} | y_{it-1}, x_{it}, z_{it}, w_i, v_i, \kappa_i) \quad (20)$$

Conditioned on the individual effects $\kappa_i = (\kappa_{1i}, \kappa_{2i})$, the observations on y_{it} are assumed to be independent. The parameter to be estimated in this model will be denoted by

$$\Theta = \left(\phi_1, \phi_2, \gamma_1, \gamma_2, \beta_1, \beta_2, \rho_{e_{it}}, \rho_{\delta_{it}}, \tau, h_0^b, h_1^b, h_2^b, h_3^b, h_0^o, h_1^o, h_2^o, h_3^o, \sigma_1, \sigma_2 \right). \text{ Hence,}$$

the likelihood function of individual i , starting from $t = 1$ and conditional on the regressors and the initial conditions, is written as,

$$L_i = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \prod_{t=1}^T \prod_{k=0}^K L_{tk} \left(y_{tk}^b, y_{tk}^o \mid y_{i0k}^b, y_{ikt-1}^b, x_{it}, \bar{x}_i, w_i, y_{i0k}^o, y_{ikt-1}^o, z_{it}, \bar{z}_i, v_i, \delta_{li}, \delta_{2i2} \right) \times \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho_{\delta_{li}\delta_{2i}}^2)}} \exp \left\{ \frac{-1}{2(1-\rho_{\delta_{li}\delta_{2i}}^2)} \left[\left(\frac{\delta_{li}}{\sigma_1} \right)^2 - 2\rho_{\delta_{li}\delta_{2i}} \left(\frac{\delta_{li}}{\sigma_1} \right) \left(\frac{\delta_{2i}}{\sigma_2} \right) + \left(\frac{\delta_{2i}}{\sigma_2} \right)^2 \right] \right\} d\delta_{li} d\delta_{2i} \quad (21)$$

Or

$$L_i = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \prod_{t=1}^T \prod_{k=0}^K \left\{ P(y_{it} = k \mid y_{i0k}^b, y_{ikt-1}^b, y_{i0k}^o, y_{ikt-1}^o, x_{it}, z_{it}, \bar{x}_i, \bar{z}_i, w_i, v_i, \Theta) \right\}^{d_{itk}} \times \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho_{\delta_{li}\delta_{2i}}^2)}} \exp \left\{ \frac{-1}{2(1-\rho_{\delta_{li}\delta_{2i}}^2)} \left[\left(\frac{\delta_{li}}{\sigma_1} \right)^2 - 2\rho_{\delta_{li}\delta_{2i}} \left(\frac{\delta_{li}}{\sigma_1} \right) \left(\frac{\delta_{2i}}{\sigma_2} \right) + \left(\frac{\delta_{2i}}{\sigma_2} \right)^2 \right] \right\} d\delta_{li} d\delta_{2i} \quad (22)$$

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

2.1. Likelihood Approximation by Gauss–Hermite Quadrature

Consider the dynamic panel ordered probit models given by equations (18) and (19). Assume that the vectors (e_{it}^b, e_{it}^o) and $(\delta_{li}, \delta_{2i})$ are independent of each other, and independently and identically distributed over time and across individuals following a normal distribution with mean

zero and covariance matrix given by $\Sigma_{\epsilon b \epsilon o}$ and $\Sigma_{\delta_{li}\delta_{2i}}$ respectively. The likelihood function of individual i , starting from $t=1$ and conditional on the regressors and the initial conditions is obtained by “integrating out” the individual effects. The likelihood function of individual i conditional on the individual effects is given in equation (22) with σ_1^2 and σ_2^2 being normalized to 1 for identification reasons. The expression of the bivariate normal distribution of δ_{li} and δ_{2i} is written as.

$$g(\delta_{li}, \delta_{2i} \mid \sigma_1^2, \sigma_2^2, \rho_{\delta_{li}\delta_{2i}}) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho_{\delta_{li}\delta_{2i}}^2)}} \times \exp \left\{ \frac{-1}{2(1-\rho_{\delta_{li}\delta_{2i}}^2)} \left[\left(\frac{\delta_{li}}{\sigma_1} \right)^2 - 2\rho_{\delta_{li}\delta_{2i}} \left(\frac{\delta_{li}}{\sigma_1} \right) \left(\frac{\delta_{2i}}{\sigma_2} \right) + \left(\frac{\delta_{2i}}{\sigma_2} \right)^2 \right] \right\} d\delta_{li} d\delta_{2i} \quad (23)$$

Hence, equation (22) can be written as,

$$L_i = \int_{-\infty}^{+\infty} \exp \left\{ \frac{-1}{2(1-\rho_{\delta_{li}\delta_{2i}}^2)} \frac{\delta_{2i}^2}{\sigma_2^2} \right\} \prod_{t=1}^T \prod_{k=0}^K M(\delta_{li}) d\delta_{2i} \quad (24)$$

where

$$M(\delta_{li}) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho_{\delta_{li}\delta_{2i}}^2)}} \int_{-\infty}^{+\infty} \exp \left\{ \frac{-1}{2(1-\rho_{\delta_{li}\delta_{2i}}^2)} \frac{\delta_{li}^2}{\sigma_1^2} \right\} \exp \left\{ \frac{1}{(1-\rho_{\delta_{li}\delta_{2i}}^2)} \rho_{\delta_{li}\delta_{2i}} \left(\frac{\delta_{li}}{\sigma_1} \right) \left(\frac{\delta_{2i}}{\sigma_2} \right) \right\} \times \prod_{t=1}^T \prod_{k=0}^K \left\{ P(y_{it} = k \mid y_{i0k}^b, y_{ikt-1}^b, y_{i0k}^o, y_{ikt-1}^o, x_{it}, z_{it}, \bar{x}_i, \bar{z}_i, w_i, v_i, \theta) \right\}^{d_{itk}} d\delta_{li} \quad (25)$$

Equation (25) can be approximated using “two-step” Gauss-Hermite quadrature, which states that

$$\int_{-\infty}^{+\infty} e^{-z^2} f(z) dz \approx \sum_{h=1}^H w_h f(a_h) \quad (26)$$

where w_h and a_h are respectively the weights and abscissas of the Gauss-Hermite quadrature and H is the total number of integration points. The larger H is the more

accurate the Gauss-Hermite quadrature. The “two-step” Gauss-Hermite quadrature consists, in the first step, in approximating equation (25) using equation (26). In the second step, a second approximation is applied to equation (25) where $M(\delta_{li})$ is replaced by its first-step Gauss-Hermite approximation.

Let $A = \phi_1 y_{it-1}^{b*} + \gamma'_1 x_{it} + \beta'_1 w_i + h_0^b + h_1^b y_{i0}^{b*} + h_2^{b*} \bar{x}_i + h_3^{b*} x_{i0}$ and $B = \phi_2 y_{it-1}^{o*} + \gamma'_2 z_{it} + \beta'_2 v_i + h_0^o + h_1^o y_{i0}^{o*} + h_2^{o*} \bar{z}_i + h_3^{o*} z_{i0}$

$$P(y_{it} = k) = \begin{cases} P(y_{it} = 0) = (1 - \Phi(A + \delta_{1i})) + \Phi(A + \delta_{1i}, -B - \delta_{2i}, -\rho_{ebeo}) \\ P(y_{it} = k) = \Phi(A + \delta_{1i}, \tau_k - B - \delta_{2i}, -\rho_{ebeo}) \\ \quad - \Phi(A + \delta_{1i}, \tau_{k-1} - B - \delta_{2i}, -\rho_{ebeo}) \\ P(y_{it} = K) = \Phi(A + \delta_{1i}, B + \delta_{2i} - \tau_{K-1}, \rho_{ebeo}) \end{cases} \quad (27)$$

Consider the first change of variable $z_{1i} = \delta_{1i} \left(\sigma_1 \sqrt{2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)} \right)^{-1}$ and $\delta_{1i} = z_{1i} \sigma_1 \sqrt{2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)}$. Differentiate z_{1i} with respect to δ_{1i} , we get,

$$\frac{dz_{1i}}{d\delta_{1i}} = \frac{1}{\sigma_1 \sqrt{2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)}} \quad \sigma_1 \sqrt{2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)} dz_{1i} = d\delta_{1i}$$

Substituting the derivatives into equation (25), we get,

$$M(\delta_{2i}) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)}} \int_{-\infty}^{+\infty} \exp\{-z_{1i}\}^2 \exp\left\{\frac{\rho_{\delta_{1i}\delta_{2i}}}{(1 - \rho_{\delta_{1i}\delta_{2i}}^2)} \left(z_{1i} \sqrt{2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)}\right) \left(\frac{\delta_{2i}}{\sigma_2}\right)\right\} \\ \times \sigma_1 \sqrt{2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)} dz_{1i} \prod_{k=0}^K \left\{ P(y_{it} = k | y_{i0k}^*, y_{ikt-1}^*, y_{i0k}^{*o}, y_{ikt-1}^{*o}, x_{it}, z_{it}, \bar{x}_i, \bar{z}_i, w_i, v_i, \theta) \right\}^{d_{itk}} \quad (28)$$

Simplifying the above equation (28), we get,

$$M(\delta_{2i}) = \frac{\sqrt{2}}{2\pi\sigma_{\delta_2}} \int_{-\infty}^{+\infty} \exp\{-z_{1i}\} \exp\left\{\frac{\rho_{\delta_{1i}\delta_{2i}}}{(1 - \rho_{\delta_{1i}\delta_{2i}}^2)} z_{1i} \sqrt{2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)} \left(\frac{\delta_{2i}}{\sigma_2}\right)\right\} \\ \times \prod_{k=0}^K \left\{ \left[\left(1 - \Phi\left(A + z_{1i} \sigma_{\delta_1} \sqrt{2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)}\right) \right) + \Phi\left(A + z_{1i} \sigma_{\delta_1} \sqrt{2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)}, -B - \delta_{2i}, -\rho_{ebeo}\right) \right] \right. \\ \times \left[\Phi\left(A + z_{1i} \sigma_1 \sqrt{2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)}, \tau_k - B - \delta_{2i}, -\rho_{ebeo}\right) - \Phi\left(A + z_{1i} \sigma_1 \sqrt{2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)}, \tau_{k-1} - B - \delta_{2i}, -\rho_{ebeo}\right) \right] \\ \left. \times \left[\Phi\left(A + z_{1i} \sigma_1 \sqrt{2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)}, B + \delta_{2i} - \tau_{K-1}, \rho_{ebeo}\right) \right] \right\}^{d_{itk}} dz_{1i} \quad (29)$$

and can be approximated using equation (26) by

$$M(\delta_{2i}) = \frac{\sqrt{2}}{2\pi\sigma_2} \sum_{h=1}^H w_h \exp\left\{\frac{\rho_{\delta_{1i}\delta_{2i}}}{(1 - \rho_{\delta_{1i}\delta_{2i}}^2)} a_h \sqrt{2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)} \left(\frac{\delta_{2i}}{\sigma_2}\right)\right\} \\ \times \prod_{k=0}^K \left\{ \left[\left(1 - \Phi\left(A + a_h \sigma_1 \sqrt{2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)}\right) \right) + \Phi\left(A + a_h \sigma_1 \sqrt{2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)}, -B - \delta_{2i}, -\rho_{ebeo}\right) \right] \right. \\ \times \left[\Phi\left(A + a_h \sigma_{\delta_1} \sqrt{2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)}, \tau_k - B - \delta_{2i}, -\rho_{ebeo}\right) - \Phi\left(A + a_h \sigma_1 \sqrt{2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)}, \tau_{k-1} - B - \delta_{2i}, -\rho_{ebeo}\right) \right] \\ \left. \times \left[\Phi\left(A + a_h \sigma_1 \sqrt{2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)}, B + \delta_{2i} - \tau_{K-1}, \rho_{ebeo}\right) \right] \right\}^{d_{itk}} \quad (30)$$

where w_h and a_h are the weights and nodes of the first-step Gauss-Hermite quadrature with H being the total number of integration points. Consider the second change of variable $z_{2i} = \delta_{2i} \left(\sigma_2 \sqrt{2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)} \right)^{-1}$ and $\delta_{2i} = z_{2i} \sigma_2 \sqrt{2(1 - \rho_{\delta_{1i}\delta_{2i}}^2)}$. Differentiate z_{2i} with respect to δ_{2i} , we get,

$$\frac{dz_{2i}}{d\delta_{2i}} = \frac{1}{\sigma_2 \sqrt{2(1-\rho_{\delta_{1i}\delta_{2i}}^2)}} \sigma_2 \sqrt{2(1-\rho_{\delta_{1i}\delta_{2i}}^2)} dz_{2i} = d\delta_{2i}$$

Replacing them in equation (30), we get,

$$L_i = \int_{-\infty}^{+\infty} \exp \left\{ \left(-\delta_{2i} \left(\sigma_2 \sqrt{2(1-\rho_{\delta_{1i}\delta_{2i}}^2)} \right)^{-1} \right)^2 \right\} \prod_{t=1}^T M(\delta_{1i}) \left(\sigma_2 \sqrt{2(1-\rho_{\delta_{1i}\delta_{2i}}^2)} \right) dz_{2i} \quad (31)$$

Substituting $z_{2i} = \delta_{2i} \left(\sigma_2 \sqrt{2(1-\rho_{\delta_{1i}\delta_{2i}}^2)} \right)^{-1}$ in equation (31), we get,

$$L_i = \int_{-\infty}^{+\infty} \exp \{-z_{2i}^2\} \prod_{t=1}^T M(\delta_{1i}) \left(\sigma_2 \sqrt{2(1-\rho_{\delta_{1i}\delta_{2i}}^2)} \right) dz_{2i} \quad (32)$$

Substituting $M(z_{1i})$ in equation (24), we get

$$\begin{aligned} L_i = & \int_{-\infty}^{+\infty} \exp \{-z_{2i}^2\} \frac{\sqrt{2}}{2\pi\sigma_2} \sum_{h=1}^H w_h \exp \left\{ \frac{\rho_{\delta_{1i}\delta_{2i}}}{(1-\rho_{\delta_{1i}\delta_{2i}}^2)} a_h \sqrt{2(1-\rho_{\delta_{1i}\delta_{2i}}^2)} \left(\frac{z_{2i} \sigma_2 \sqrt{2(1-\rho_{\delta_{1i}\delta_{2i}}^2)}}{\sigma_2} \right) \right\} \\ & \times \left(\sigma_2 \sqrt{2(1-\rho_{\delta_{1i}\delta_{2i}}^2)} \right) dz_{2i} \prod_{t=1}^T \prod_{k=0}^K \left\{ \left[\left(1 - \Phi \left(A + a_h \sigma_1 \sqrt{2(1-\rho_{\delta_{1i}\delta_{2i}}^2)} \right) \right) + \Phi \left(A + a_h \sigma_1 \sqrt{2(1-\rho_{\delta_{1i}\delta_{2i}}^2)}, -B - \delta_{2i}, -\rho_{ebeo} \right) \right] \right. \\ & \times \left[\Phi \left(A + a_h \sigma_1 \sqrt{2(1-\rho_{\delta_{1i}\delta_{2i}}^2)}, \tau_k - B - \delta_{2i}, -\rho_{ebeo} \right) - \Phi \left(A + a_h \sigma_1 \sqrt{2(1-\rho_{\delta_{1i}\delta_{2i}}^2)}, \tau_{k-1} - B_i - \delta_{2i}, -\rho_{ebeo} \right) \right] \\ & \times \left. \left[\Phi \left(A + a_h \sigma_1 \sqrt{2(1-\rho_{\delta_{1i}\delta_{2i}}^2)}, B + \delta_{2i} - \tau_{K-1}, \rho_{ebeo} \right) \right] \right\}^{d_{ik}} \end{aligned} \quad (33)$$

Let

$$\begin{aligned} u_{1,qh} &= \phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1 w_i + h_0^b + h_1^b y_{i0}^b + h_2^{b'} \bar{x}_i + h_3^{b'} x_{i0} + a_h \sigma_1 \sqrt{2(1-\rho_{\delta_{1i}\delta_{2i}}^2)} \\ u_{1,qhk} &= \tau_k - \phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2 v_i - h_0^o - h_1^o y_{i0}^o - h_2^{o'} z_i - h_3^{o'} z_{i0} - a_q \sigma_2 \sqrt{2(1-\rho_{\delta_{1i}\delta_{2i}}^2)} \\ u_{1,qh0} &= \tau_0 - \phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2 v_i - h_0^o - h_1^o y_{i0}^o - h_2^{o'} \bar{z}_i - h_3^{o'} z_{i0} - a_q \sigma_2 \sqrt{2(1-\rho_{\delta_{1i}\delta_{2i}}^2)} \\ u_{1,qhk-1} &= \tau_{k-1} - \phi_2 y_{it-1}^o - \gamma_2' z_{it} - \beta_2 v_i - h_0^o - h_1^o y_{i0}^o - h_2^{o'} \bar{z}_i - h_3^{o'} z_{i0} - a_q \sigma_2 \sqrt{2(1-\rho_{\delta_{1i}\delta_{2i}}^2)} \\ u_{1,qhK} &= \phi_2 y_{it-1}^o + \gamma_2' z_{it} + \beta_2 v_i + h_0^o + h_1^o y_{i0}^o + h_2^{o'} \bar{z}_i + h_3^{o'} z_{i0} + a_q \sigma_2 \sqrt{2(1-\rho_{\delta_{1i}\delta_{2i}}^2)} - \tau_{K-1} \end{aligned}$$

Simplifying equation (33), we get

$$\begin{aligned} L_i = & \int_{-\infty}^{+\infty} \exp \{-z_{2i}^2\} \prod_{t=1}^T \pi^{-1} \sqrt{(1-\rho_{\delta_{1i}\delta_{2i}}^2)} \sum_{h=1}^H w_h \exp \{2\rho_{\delta_{1i}\delta_{2i}} z_{2i} a_h\} \\ & \times \prod_{t=1}^T \prod_{k=0}^K \left\{ \left[\left(1 - \Phi(u_{1,qh}) \right) + \Phi(u_{1,qh}, u_{1,qh0}, -\rho_{ebeo}) \right] \right. \\ & \times \left. \left[\Phi(u_{1,qh}, u_{1,qhk}, -\rho_{ebeo}) - \Phi(u_{1,qh}, u_{1,qhk-1}, -\rho_{ebeo}) \right] \left[\Phi(u_{1,qh}, u_{1,qK}, \rho_{ebeo}) \right] \right\}^{d_{ik}} dz_{2i} \end{aligned} \quad (34)$$

and can be approximated using equation (26), we get

$$L_i = \pi^{-1} \sqrt{(1 - \rho_{\delta_{1i}\delta_{2i}}^2)} \sum_{q=1}^Q \sum_{h=1}^H w_q w_h \exp\{2\rho_{\delta_{1i}\delta_{2i}} a_q a_h\} \times \prod_{t=1}^T \prod_{k=0}^K \left[\left((1 - \Phi(u_{1,qh})) + \Phi(u_{1,qh}, u_{1,qh0}, -\rho_{e_{beo}}) \right) \right] \times \left[\Phi(u_{1,qh}, u_{1,qhk}, -\rho_{e_{beo}}) - \Phi(u_{1,qh}, u_{1,qhk-1}, -\rho_{e_{beo}}) \right] \left[\Phi(u_{1,qh}, u_{1,qK}, \rho_{e_{beo}}) \right] \}^{d_{ik}} \quad (35)$$

where w_q and a_q are the weights and nodes of the second-step Gauss-Hermite quadrature with H being the total number of integration points. The product over i of the approximate likelihood function can be maximized

using Newton Raphson Method procedures to obtain estimates of the parameters. Finally, as the individuals are independent, the log-likelihood function should be expressed as:

$$l(\Theta, y) = \sum_{i=1}^N \left\{ \log \left[\pi^{-1} \sqrt{(1 - \rho_{\delta_{1i}\delta_{2i}}^2)} \sum_{q=1}^Q \sum_{h=1}^H w_q w_h \exp\{2\rho_{\delta_{1i}\delta_{2i}} a_q a_h\} \times \prod_{t=1}^T \prod_{k=0}^K \left[\left((1 - \Phi(u_{1,qh})) + \Phi(u_{1,qh}, u_{1,qh0}, -\rho_{e_{beo}}) \right) \right] \right] \right\} \times \left[\Phi(u_{1,qh}, u_{1,qhk}, -\rho_{e_{beo}}) - \Phi(u_{1,qh}, u_{1,qhk-1}, -\rho_{e_{beo}}) \right] \left[\Phi(u_{1,qh}, u_{1,qK}, \rho_{e_{beo}}) \right] \}^{d_{ik}} \quad (36)$$

Note that we present the auxiliary conditional distribution of κ_{1i} with a constant h_0^b . Thus, the constant in the structural equation have been dropped. Note that we present the auxiliary conditional distribution of κ_{2i} with a constant h_0^o .

Thus, the constant in the structural equation have been dropped.

When the error terms are independent, the likelihood function is given by,

$$\ell(\Theta, y) = \sum_{i=1}^N \log \left\{ \pi^{-1} \sqrt{(1 - \rho_{\delta_{1i}\delta_{2i}}^2)} \sum_{q=1}^Q \sum_{h=1}^H w_q w_h \exp\{2\rho_{\delta_{1i}\delta_{2i}} a_q a_h\} \times \prod_{t=1}^T \prod_{k=0}^K \left[\left((1 - \Phi(u_{1,qh})) + \Phi(u_{1,qh}, u_{1,qh0}, -\rho_{e_{beo}}) \right) \right] \right\} \times \left[\Phi(u_{1,qh}, u_{1,qhk}, -\rho_{e_{beo}}) - \Phi(u_{1,qh}, u_{1,qhk-1}, -\rho_{e_{beo}}) \right] \left[\Phi(u_{1,qh}, u_{1,qK}, \rho_{e_{beo}}) \right] \}^{d_{ik}} \quad (37)$$

At each evaluation of the likelihood function, it is necessary to compute $N \times Q \times H$ cumulative density functions of the bivariate normal variables with this

two-step quadrature. To find the approximate values of the MLE's of Θ , we applied the Newton-Raphson algorithm. It is given by,

$$\begin{bmatrix} \hat{\psi}^{(t+1)} \\ \hat{\lambda}^{(t+1)} \end{bmatrix} = \begin{bmatrix} \hat{\psi}^{(t)} \\ \hat{\lambda}^{(t)} \end{bmatrix} + \begin{bmatrix} -\frac{\partial^2 \ell(\psi, \lambda, y)}{\partial \psi \partial \psi^t} & -\frac{\partial^2 \ell(\psi, \lambda, y)}{\partial \psi \partial \lambda} \\ -\frac{\partial^2 \ell(\psi, \lambda, y)}{\partial \lambda \partial \psi} & -\frac{\partial^2 \ell(\psi, \lambda, y)}{\partial \lambda \partial \lambda^t} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial \ell(\psi, \lambda, y)}{\partial \psi} \\ \frac{\partial \ell(\psi, \lambda, y)}{\partial \lambda} \end{bmatrix} \quad (38)$$

where $t=0, 1, 2, \dots$ is the iteration number. The maximum of the log-likelihood function can be found by applying

$$\Theta^{t+1} = \Theta^t + H(\Theta^t)^{-1} G(\Theta^t) \quad (39)$$

until convergence is reached, where $G(\Theta^t)$ is the gradient evaluated at Θ^t and $H(\Theta^t)$ is the Hessian matrix evaluated at Θ^t . This means that starting values Θ^0 are needed. The algorithm is said to have converged when the log likelihood changes by a less than a small constant $\varepsilon > 0$, that is, $|\Theta^{t+1} - \Theta^t| < \varepsilon$ where, for instance, $\varepsilon = 10^{-5}$. According to Efron and Hinkley [19], to compute the standard errors of the parameter estimates we could use the observed information

matrix, i.e. minus the second-order derivative of the log-likelihood function. The parameter estimates are asymptotically normally distributed: $\hat{\Theta}_{ML} \sim N(\Theta_0, H^{-1})$.

2.2. Model Selection

We assess the statistical fit of the different models using Akaike Information Criteria (AIC) for model selection. Formally,

$$AIC = -2 \ln L + 2q \quad (40)$$

where q represents the number of parameters in each specification and $\ln L$ the maximized log-likelihood function. The preferred model is the one with the smallest value of AIC.

2.3. Average Partial Effects

To provide an indication of the magnitude of the associations between ordinal observation and the regressors, we present average partial effects (APEs). The average partial effects is computed by scaling the coefficient vector. The average partial effects provide an indication of the magnitude of the associations between the observed participation and consumption levels and the regressors. In this case 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 $\gamma_a = \gamma / \sqrt{1 + \sigma^2}$. The sign

of the APE has a clear qualitative interpretation, with a positive sign implying 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 four 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:

$$APE_i(\lambda_{\omega_i}, y_{it}^{b*} = 1) = \frac{\beta}{NT} \sum_{i=1}^N \sum_{t=1}^T \left[\Phi(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + 1w_i + h_0^b + h_1^b y_{i0}^b + h_2^b \bar{x}_i) - \Phi(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + 0w_i + h_0^b + h_1^b y_{i0}^b + h_2^b \bar{x}_i) \right] \quad (41)$$

For continuous explanatory variables, the marginal effect is given by the partial derivative of the probability of interest with respect to ω_i , Wooldridge [20] 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 marginal effect on the probability of participation is given by;

$$APE_i(\lambda_{\omega_i}, y_{it}^{b*} = 1) = \frac{\lambda_{\omega_i}}{NT} \sum_{i=1}^N \sum_{t=1}^T \varphi(\phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_0^b + h_1^b y_{i0}^b + h_2^b \bar{x}_i) \quad (42)$$

where $\varphi(\cdot)$ is the probability density function (p.d.f.) of the standard normal distribution and λ_{ω_i} is the coefficient in the inflation part associated with variable ω_i . In terms of the zeros category, the effect on the probability of non-participation (zero inflation) is

$$APE_i(\lambda_{\omega_i}, y_{it}^{b*} = 0) = \frac{\lambda_{\omega_i}}{NT} \sum_{i=1}^N \sum_{t=1}^T -\varphi(u_{1,qh}) \quad (43)$$

while

$$APE_i(\lambda_{\omega_i}, \beta_{\omega_i}, y_{it}^{b*} = 1, y_{it}^{o*} = 0) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left\{ \left[\Phi\left(\frac{-u_{1,qh0} + \rho_{eboo} u_{1,qh}}{\sqrt{1 - \rho_{eboo}^2}}\right) \right] \varphi(u_{1,qh}) \lambda_{\omega_i} - \left[\Phi\left(\frac{u_{1,qh} - \rho_{eboo} u_{1,qh0}}{\sqrt{1 - \rho_{eboo}^2}}\right) \right] \varphi(-u_{1,qh0}) \beta_{\omega_i} \right\} \quad (44)$$

represents the marginal effect on the probability of zero consumption.

The total marginal effect on the probability of observing zero consumption is obtained as a sum of the marginal effects in (54) and (55); that is,

$$APE_i(y_{it}^{o*} = 0) = APE_i(y_{it}^{b*} = 0) + APE_i(y_{it}^{b*} = 1, y_{it}^{o*} = 0) \quad (45)$$

Let $u_{1,qhk} = \tau_k - W$ $u_{1,qhk-1} = \tau_{k-1} - W$ and $u_{1,qhK} = W - \tau_{K-1}$

$$APE_i(\lambda_{\omega_i}, \beta_{\omega_i}, y_{it}^{b*} = 1, y_{it}^{o*} = 1) = \sum_{i=1}^N \sum_{t=1}^T \left[\Phi\left(\frac{(\tau_2 - W) + \rho_{eboo} u_{1,qh}}{\sqrt{1 - \rho_{eboo}^2}}\right) - \Phi\left(\frac{-W + \rho_{eboo} u_{1,qh}}{\sqrt{1 - \rho_{eboo}^2}}\right) \right] \varphi(u_{1,qh}) \lambda_{\omega_i} - \left[\Phi\left(\frac{u_{1,qh} + \rho_{eboo} (\tau_2 - W)}{\sqrt{1 - \rho_{eboo}^2}}\right) \varphi(\tau_2 - W) - \Phi\left(\frac{u_{1,qh} - \rho_{eboo} W}{\sqrt{1 - \rho_{eboo}^2}}\right) \varphi(-W) \right] \beta_{\omega_i} \quad (46)$$

$$APE_i(\lambda_{\omega_i}, \beta_{\omega_i}, y_{it}^{b*} = 1, y_{it}^{o*} = 2) = \sum_{i=1}^N \sum_{t=1}^T \left[\Phi \left(\frac{(\tau_3 - W) + \rho_{ebeo} u_{1,qh}}{\sqrt{1 - \rho_{ebeo}^2}} \right) - \Phi \left(\frac{\tau_2 - W + \rho_{ebeo} u_{1,qh}}{\sqrt{1 - \rho_{ebeo}^2}} \right) \right] \varphi(u_{1,qh}) \lambda_{\omega_i} \\ - \left[\Phi \left(\frac{u_{1,qh} + \rho_{ebeo} (\tau_3 - W)}{\sqrt{1 - \rho_{ebeo}^2}} \right) \varphi(\tau_3 - W) - \Phi \left(\frac{u_{1,qh} + \rho_{ebeo} (\tau_2 - W)}{\sqrt{1 - \rho_{ebeo}^2}} \right) \varphi(\tau_2 - W) \right] \beta_{\omega_i} \quad (47)$$

$$APE_i(\lambda_{\omega_i}, \beta_{\omega_i}, y_{it}^{b*} = 1, y_{it}^{o*} = k) = \sum_{i=1}^N \sum_{t=1}^T \left[\Phi \left(\frac{(\tau_k - W) + \rho_{ebeo} u_{1,qh}}{\sqrt{1 - \rho_{ebeo}^2}} \right) - \Phi \left(\frac{\tau_{k-1} - W + \rho_{ebeo} u_{1,qh}}{\sqrt{1 - \rho_{ebeo}^2}} \right) \right] \varphi(u_{1,qh}) \lambda_{\omega_i} \\ - \left[\Phi \left(\frac{u_{1,qh} + \rho_{ebeo} (\tau_k - W)}{\sqrt{1 - \rho_{ebeo}^2}} \right) \varphi(\tau_k - W) - \Phi \left(\frac{u_{1,qh} + \rho_{ebeo} (\tau_{k-1} - W)}{\sqrt{1 - \rho_{ebeo}^2}} \right) \varphi(\tau_{k-1} - W) \right] \beta_{\omega_i} \quad (48)$$

$$APE_i(\lambda_{\omega_i}, \beta_{\omega_i}, y_{it}^{b*} = 1, y_{it}^{o*} = J) = \sum_{i=1}^N \sum_{t=1}^T \left[\Phi \left(1 - \frac{(\tau_K - W) + \rho_{ebeo} u_{1,qh}}{\sqrt{1 - \rho_{ebeo}^2}} \right) \right] \varphi(u_{1,qh}) \lambda_{\omega_i} + \left[\Phi \left(\frac{u_{1,qh} + \rho_{ebeo} (\tau_K - W)}{\sqrt{1 - \rho_{ebeo}^2}} \right) \varphi(\tau_K - W) \right] \beta_{\omega_i} \quad (49)$$

For each choice probability, the marginal effect consists of two parts: the effects coming from the inflation sub-model and the component attributable to the probit part of the model.

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

$$Var(\widehat{APE}) = \nabla_{\Theta} APE_{k,i,t} Var(\Theta) \nabla_{\Theta} APE'_{k,i,t} \quad (50)$$

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 $\phi_2 = 0.5$, $\gamma_{21} = 0.6$, $\gamma_{22} = 1.5$, $\beta_{21} = 1.0$, $\sigma_2 = 1.0$. The inter-period correlation was $\sigma_2^2 / (\sigma_2^2 + 1) = 0.5$. The covariance matrix was given by

$$\Sigma_{\delta_i, \delta_{2i}} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \text{ and the correlation matrix was given by } \Sigma_{ebeo} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}. \text{ The number of replication (R) was 200.}$$

We maximized the log-likelihood function using Newton-Raphson algorithm. This algorithm requires starting values. To have a warm start for finding the global maximum of the log-likelihood function, the study first estimated the dynamic panel binary probit model and dynamic panel ordered probit model. We used the parameter estimates of these models as starting values for the Newton-Raphson algorithm. Using these starting values, convergence was reached quite fast. The algorithm was said to have converged when the log-likelihood changed by a less than a small constant $\varepsilon < 10^{-5}$. Lesaffre and Spiessens [21] showed that a number of 10 nodes and weights is often sufficient and differences by further increasing nodes and weights are only minimal.

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:

$$\bar{\theta}_k = \frac{1}{R} \sum_{r=1}^R \hat{\theta}_k^r. \text{ To assess the accuracy of the parameter,}$$

which incorporates both bias and variability, Root Mean Square Error (RMSE) were calculated for each parameter. RMSE is always non-negative, and a value of 0 (almost never achieved in practice) would indicate a perfect fit to the data. In general, a lower RMSE is better than a higher one. The Root Mean Square Error was given by,

$$\bar{\theta}_k = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\theta}_k^r - \theta)^2}.$$

The Table 1 below indicates the results of the three models from the simulation.

Table 1. Maximum likelihood estimates based on $n=750$, $T=10$ and 10-Point Gauss Hermite quadrature for Correlated Zero Inflated Dynamic Panel Ordered Probit (ZIDPOPC) in simulated data.

MODELS		DPOP			ZIDPOPI			ZIDPOPC		
Par	True	Ests	Stderr	P val	Ests	Stderr	P val	Ests	Stderr	P val
h_0^b					-0.623	0.088	0.000	-0.681	0.085	0.000
ϕ_1	0.5				0.533	0.062	0.000	0.591	0.062	0.000
γ_{11}	-1.0				-0.934	0.041	0.000	-0.978	0.040	0.000
β_1	-1.0				-0.643	0.089	0.000	-0.658	0.088	0.000
h_1^b					1.175	0.103	0.000	1.144	0.103	0.000
h_{2m}^b					-0.072	0.148	0.627	0.067	0.147	0.650
h_{2i}^b					0.239	0.049	0.000	0.259	0.048	0.000
σ_1	1.0				0.758	0.077	0.000	0.771	0.072	0.000
h_0^o		-1.103	0.055	0.000	-0.487	0.096	0.000	-0.649	0.078	0.000
ϕ_2	0.5	0.241	0.025	0.000	0.581	0.043	0.000	0.578	0.040	0.000
γ_{21}	0.4	0.148	0.017	0.000	0.412	0.035	0.000	0.400	0.028	0.000
γ_{22}	1.0	0.406	0.018	0.000	1.003	0.061	0.000	0.997	0.036	0.000
β_2	0.7	-0.456	0.062	0.000	0.546	0.104	0.000	0.367	0.085	0.000
h_1^o		0.562	0.046	0.650	0.847	0.089	0.000	0.867	0.065	0.000
h_{2m1}^o		0.003	0.105	0.978	0.014	0.143	0.927	-0.051	0.135	0.708
h_{2m2}^o		-0.074	0.107	0.488	0.025	0.145	0.865	0.075	0.137	0.585
h_{2i1}^o		-0.102	0.033	0.002	-0.149	0.043	0.001	-0.137	0.040	0.001
h_{2i1}^o		-0.208	0.035	0.000	-0.332	0.053	0.000	-0.372	0.048	0.000
σ_2	1.0	0.948	0.042	0.000	0.670	0.108	0.000	0.674	0.074	0.000
$\rho_{e\beta eo}$	0.5							0.462	0.093	0.000
$\rho_{\delta_{1i}, \delta_{2i}}$	0.5				0.337	0.236	0.154	0.403	0.105	0.000
τ_1	2.6	1.119	0.025	0.000	2.533	0.163	0.000	2.536	0.083	0.000
τ_2	4.2	2.084	0.043	0.000	4.154	0.143	0.000	4.169	0.071	0.000
AIC		11501.730			9347.912			8664.670		

3.2. Results for Maximum Likelihood Estimation Using Gauss Hermite Quadrature Approximation for Simulated Data

Table 1 shows the parameters, true values, estimates, standard errors, p values and AIC values for DPOP, ZIDPOPI and ZIDPOPC models. All the parameters whose p values are less than 0.05 are significant at 5%. The initial observations in both participation decision h_1^b and consumption levels h_1^o were significant at 1% in the three models. The correlation between the error terms in ZIDPOPC model was significant at 1% implying that the factors affecting the participation decision are the same as the one affecting the consumption levels. The correlation between the unobserved individual effects in ZIDPOPC model was not significant at 5% implying that the factors affecting the unobserved individual effects in participation decision are not the same as the one affecting the unobserved

individual effects at consumption levels. The variance of the individual effects in participation decision was 0.771. This indicated that 37.28% of the latent error variance is attributable to unobserved heterogeneity, as measured by the intra-unit correlation coefficient in smoking decision. The variance of the individual effects for the decision on consumption levels was 0.674. Approximately 31.24% of the latent error variance is attributable to unobserved heterogeneity, as measured by the intra-unit correlation coefficient at consumption levels. The Akaike Information Criteria (AIC) was given as a measure of goodness-of-fit. The model with the smallest AIC was considered to fit the data better than the rest. The ZIDPOPC model clearly fitted the data better than ZIDPOPI and DPOP models. The ZIDPOPI model clearly fitted the data better than 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			0.533	0.058	0.591	0.102
γ_{11}	-1.0			-0.934	0.074	-0.978	0.049
β_1	-1.0			-0.643	0.362	-0.658	0.357
σ_1	1.0			0.758	0.242	0.771	0.238
ϕ_2	0.5	0.241	0.260	0.581	0.082	0.578	0.080
γ_{21}	0.4	0.148	0.252	0.412	0.040	0.400	0.015
γ_{22}	1.0	0.406	0.594	1.003	0.011	0.997	0.023
β_2	0.7	-0.456	1.158	0.546	0.197	0.367	0.345
σ_2	1.0	0.948	0.062	0.670	0.332	0.674	0.329
ρ_{ebeo}	0.5					0.462	0.080
$\rho_{\delta_{1i}, \delta_{2i}}$	0.5			0.337	0.176	0.403	0.131
τ_1	2.6	1.119	1.482	2.533	0.089	2.536	0.092
τ_2	4.2	2.084	2.117	4.154	0.059	4.169	0.082

Table 2 shows the parameters, true values, estimates and RMSE for the three models when $n=750$ and $T=10$. Although the ZIDPOPI and ZIDPOPC models had more parameters than DPOP model, for the comparison purpose, comments were restricted to only the parameters that are common to all models, that is, ϕ_2 , γ_{21} , γ_{22} , β_2 , σ_2 , τ_1 and τ_2 . 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's estimates.

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

Table 3. Maximum likelihood estimates based on $n=350$ and $T=10$ for DPOP, ZIDPOPI and ZIDPOPC models.

MODELS		DPOP		ZIDPOPI		ZIDPOPC	
Par	True	Ests	Stderr	Ests	Stderr	Ests	Stderr
ϕ_1	0.5			0.542	0.092	0.612	0.094
γ_{11}	-1.0			-0.953	0.060	-0.975	0.061
β_1	-1.0			-0.745	0.137	-0.714	0.137
σ_1	1.0			0.780	0.106	0.810	0.107
ϕ_2	0.5	0.236	0.036	0.544	0.060	0.566	0.059
γ_{21}	0.4	0.157	0.024	0.416	0.730	0.369	0.040
γ_{22}	1.0	0.390	0.026	1.001	0.059	0.972	0.052
β_2	0.7	-0.480	0.091	0.586	0.145	0.387	0.128
σ_2	1.0	0.933	0.061	0.730	0.116	0.676	0.109
ρ_{ebeo}	0.5					0.476	0.141
$\rho_{\delta_{1i}, \delta_{2i}}$	0.5			0.264	0.200	0.391	0.157
τ_1	2.6	1.110	0.037	2.567	0.150	2.558	0.121
τ_2	4.2	2.059	0.063	4.150	0.131	4.138	0.107

Table 4. Maximum likelihood estimates based on $n=750$ and $T=10$ for DPOP, ZIDPOPI and ZIDPOPC models.

MODELS		DPOP		ZIDPOPI		ZIDPOPC	
Par	True	Ests	Stderr	Ests	Stderr	Ests	Stderr
ϕ_1	0.5			0.533	0.062	0.591	0.062
γ_{11}	-1.0			-0.934	0.041	-0.978	0.040
β_1	-1.0			-0.643	0.089	-0.658	0.088

MODELS		DPOP		ZIDPOPI		ZIDPOPC	
Par	True	Ests	Stderr	Ests	Stderr	Ests	Stderr
σ_1	1.0			0.758	0.077	0.771	0.072
ϕ_2	0.5	0.241	0.025	0.581	0.043	0.578	0.040
γ_{21}	0.4	0.148	0.017	0.412	0.035	0.400	0.028
γ_{22}	1.0	0.406	0.018	1.003	0.061	0.997	0.036
β_2	0.7	-0.456	0.062	0.546	0.104	0.367	0.085
σ_2	1.0	0.948	0.042	0.670	0.108	0.674	0.074
ρ_{ebeo}	0.5					0.462	0.093
$\rho_{\delta_{1i}\delta_{2i}}$	0.5			0.337	0.236	0.403	0.105
τ_1	2.6	1.119	0.025	2.533	0.163	2.536	0.083
τ_2	4.2	2.084	0.043	4.154	0.143	4.169	0.071

3.4. Assessing the Consistency of the Estimators of in Models

The study compared the 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 0.612 and when $n=750$, ϕ_1 was 0.591 for ZIDPOPC model. 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 paper was the National Longitudinal Survey of Youth 1997 (NLSY97). 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, we 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 ordered values $y_{it} = 0, 1, 2, 3$, stands roughly for zero, low, moderate and high levels of tobacco consumption.

3.6. Results for Maximum Likelihood Estimation Using Gauss Hermite Quadrature Approximation for Real Data

The Table 5 below gives the results of the three models based on real data.

Table 5. Maximum likelihood estimates based on Gauss Hermite quadrature Approximation for $n=2500$ and $T=8$ for Correlated Zero Inflated Dynamic Panel Ordered Probit Model in Smoking Data.

MODELS		DPOP		ZIDPOPI			ZIDPOPC		
Param	Ests	Stderror	P val	Ests	Stderr	P val	Ests	Stderr	P val
ϕ				1.340	0.083	0.000	1.406	0.095	0.000
Age				-0.036	0.049	0.460	0.026	0.033	0.436
Edu				0.062	0.063	0.324	0.055	0.031	0.080
Gender				0.087	0.111	0.431	0.101	0.068	0.141
Race				-0.399	0.152	0.009	-0.325	0.091	0.000
h_1^b				1.821	0.172	0.000	1.584	0.113	0.000
σ_1				1.768	0.142	0.000	0.998	0.059	0.969

MODELS		DPOP		ZIDPOPI			ZIDPOPC		
Param	Ests	Stderror	P val	Ests	Stderr	P val	Ests	Stderr	P val
ϕ_2	0.576	0.016	0.000	0.452	0.027	0.000	0.419	0.026	0.000
Age	0.064	0.015	0.000	0.102	0.022	0.000	0.090	0.024	0.000
Edu	0.030	0.013	0.027	0.037	0.016	0.023	0.014	0.016	0.395
Gender	0.159	0.039	0.000	0.235	0.049	0.000	0.271	0.064	0.000
h_2^0	0.876	0.029	0.000	0.479	0.039	0.000	0.603	0.043	0.000
σ_2	1.382	0.030	0.000	0.690	0.049	0.000	0.954	0.051	0.354
$\rho_{e_{beo}}$							0.020	0.088	0.818
$\rho_{\delta_{1i}\delta_{2i}}$				0.735	0.075	0.000	0.212	0.072	0.003
τ_1	1.331	0.019	0.000	1.496	0.031	0.000	2.124	0.133	0.000
τ_2	2.421	0.028	0.000	2.865	0.032	0.000	3.305	0.131	0.205
AIC		25855.38			25338.1088			24777.4488	

Table 5 shows the parameters, estimates, standard errors, p values and AIC values. The variables whose p values are less than 0.05 are significant at 5%. In case of ZIDPOPC model, state dependence, race and initial observation from the decision to smoke or not are significant at 5% while state dependence, age, gender, initial observation, correlation between individual effects and the first cut point from decision on the number of cigarettes smoked are significant at 5%. The variance of the individual effects in participation decision was 0.998. This indicated that 49.90% of the latent error variance is attributable to unobserved heterogeneity, as measured by the intra-unit correlation coefficient in smoking decision. The variance of the individual effects for the decision on consumption levels was 0.954. Approximately 47.65% of the latent error variance is attributable to unobserved heterogeneity, as measured by the intra-unit correlation coefficient at consumption levels.

The correlation coefficient between the error terms was not significant at 5%. This implied that the variables affecting the participation decision are different from the one affecting consumption levels. The correlation coefficient between the individual effects was significant at 1%. This implied that the variables affecting the individual effects at participation decision were the same as the one affecting individual effects

at consumption levels.

The estimated coefficients for initial period smoking decision was significant at 1%, which implies a positive association between the initial period smoking observation and unobserved latent smoking. Therefore, this indicates that it is necessary to control for smoking decision at the beginning of observations. The estimated coefficients for initial period decision on the number of cigarettes smoked observation was significant at 1%, the decision on consumption levels, which implied a positive association between the initial period consumption levels observation and unobserved latent consumption levels. Therefore, this indicates that it is necessary to control for reported smoking at the beginning of the observations.

The Akaike Information Criteria (AIC) was given as a measure of goodness-of-fit. The model with the smallest AIC is considered to fit the data better than the rest. The ZIDPOPC model clearly fitted the data better than the ZIDPOPI and DPOP models. The ZIDPOPI model clearly fitted the data better than the DPOP model.

3.7. Average Partial Effects

The Table 6 below gives the results of the Average partial effects for the three models based on real data.

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

Param		Ests	Stderror	t_value	P_value
ϕ_1	$y_{it}^b = 0$	0.328	0.031	10.524	0.000
	$y_{it}^b = 1$	0.324	0.031	10.524	0.000
Age	$y_{it}^b = 0$	-0.003	0.000	-30.423	0.000
	$y_{it}^b = 1$	0.004	0.000	30.423	0.000
Edu	$y_{it}^b = 0$	-0.007	0.000	-32.041	0.000
	$y_{it}^b = 1$	0.008	0.000	32.041	0.000
Gender	$y_{it}^b = 0$	0.018	0.001	14.633	0.000
	$y_{it}^b = 1$	0.020	0.001	14.633	0.000
Race	$y_{it}^b = 0$	-0.057	0.005	-10.941	0.000
	$y_{it}^b = 1$	-0.065	0.006	-10.941	0.000

Param		Ests	Stderror	t value	P value
ϕ_2	$y_{it}^0 = 0$	-0.029	0.001	-38.730	0.000
	$y_{it}^0 = 1$	0.025	0.001	38.730	0.000
	$y_{it}^0 = 2$	0.040	0.001	38.730	0.000
	$y_{it}^0 = 3$	0.017	0.000	38.730	0.000
Age	$y_{it}^0 = 0$	-0.008	0.000	-41.754	0.000
	$y_{it}^0 = 1$	0.000	0.000	-41.754	0.000
	$y_{it}^0 = 2$	0.012	0.000	41.754	0.000
	$y_{it}^0 = 3$	0.020	0.000	41.754	0.000
Edu	$y_{it}^0 = 0$	-0.005	0.000	-61.881	0.000
	$y_{it}^0 = 1$	0.007	0.000	61.881	0.000
	$y_{it}^0 = 2$	0.004	0.000	61.881	0.000
	$y_{it}^0 = 3$	0.004	0.000	61.881	0.000
Gender	$y_{it}^0 = 0$	0.007	0.001	8.868	0.000
	$y_{it}^0 = 1$	0.436	0.028	15.518	0.000
	$y_{it}^0 = 2$	0.194	0.013	15.518	0.000
	$y_{it}^0 = 3$	0.024	0.002	15.518	0.000
ϕ_2	$y_{it}^0 = 0$	-0.029	0.015	-1.935	0.053
Age	$y_{it}^0 = 0$	-0.005	0.000	-38.730	0.000
Edu	$y_{it}^0 = 0$	0.002	0.000	41.754	0.000
Gender	$y_{it}^0 = 0$	-0.012	0.000	-61.881	0.000

Table 6 indicates the estimated average partial effects, standard errors and p values.

It summarizes the APEs of the time invariant variables, time variant variables and state dependence at participation and consumption levels. All the covariates whose p values are less than 0.05 are statistically significant at 5%. The covariates with positive sign implying a positive association with the participation or consumption at all levels and vice versa. The average partial effects for the state dependence in non-participation decision and participation decision were 0.328 and 0.324 respectively. This indicated that as one move from the previous year to the next, the probability of non-smoking and smoking will increase by 32.8% and 32.4% respectively. The average partial effects for the age in non-participation decision and participation decision were -0.003 and 0.004 respectively. This indicated that one-unit increase in age will produce a 0.3% decrease in the probability of non-smoking for an otherwise “average” individual while one unit increase in age will produce a 0.4% increase 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 -0.029, 0.025, 0.040 and 0.017, 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 -2.9%, 2.5%, 4.0% and 1.7% respectively. The average partial effects for the age, at various consumption levels, that is, zero, low, moderate and high consumption levels were -0.008, 0.000, 0.012 and 0.020, respectively. This indicated that one-unit increase in age will produce -0.8%, 0.0%, 1.2% and 2.0% 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 -0.029, -0.005, 0.002 and -0.012, 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 -2.9%, -0.5%, 0.2% and -1.2% respectively.

4. Conclusion

The main objective of this study was to develop Zero inflated dynamic panel ordered probit model with both correlated and uncorrelated error terms and compared them with Dynamic panel ordered probit model. The Akaike information criteria from the simulation studies showed that ZIDPOPC model fitted the data better than ZIDPOPI and DPOP models while ZIDPOPI model fitted the data better than DPOP model. ZIDPOPI and ZIDPOPC estimators had a smaller RMSE compared to DPOP estimators. This showed

that ZIDPOPC and ZIDPOPI models' estimators were more accurate than DPOP's estimators. The estimates of the ZIDPOPI and ZIDPOPC models tended to the true parameter values as n tended to infinity. This indicated that ZIDPOPI and ZIDPOPC 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 estimated coefficients for initial period participation observations were statistically significant at 1%, which implied a positive correlation between the initial period participation observation and unobserved latent participation. Similarly, the estimated coefficients for initial period consumption observations were statistically significant at 1%, which implied a positive correlation between the initial period consumption observation and unobserved latent consumption. This implied that it is necessary to control for participation decision and consumption levels at the beginning of the observations. The correlation between the unobserved individual effects in ZIDPOPC model was not significant at 5% implying that the factors affecting the participation decision are not the same as the one affecting the consumption levels.

References

- [1] Gurmu, S. and Dagne, G. A. (2009). 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.
- [2] Harris, M. N. and Zhao, X. (2007). A Zero-inflated Ordered Probit Model, with an Application to Modelling Tobacco Consumption, *Journal of Econometrics*, 141, 1073-1099.
- [3] Downward, P., Lera-Lopez, F., Rasciute, S. (2011). The zero-inflated ordered Probit approach to modelling sports participation. *Economic Modelling*, 28 (6), 2469-2477.
- [4] Oh, M, Oh, H. T and Park, S. (2012). Bayesian Analysis of Korean Alcohol Consumption Data Using a Zero-Inflated Ordered Probit Model. *Korean Journal of Applied Statistics*, Vol. 25, pp. 363-376.
- [5] Bagozzi, B. E., Hill, D. W., Moore, W. H. and Mukherjee, B. (2015). Modeling Two Types of Peace The Zero-inflated Ordered Probit Model in Conflict Research. *Journal of Conflict Resolution*, 59 (4), 728-752.
- [6] Jinzhong, W., Wenji, F., and Wencheng, W. (2020) A Zero-Inflated Ordered Probit Model to Analyze Hazmat Truck Drivers' Violation Behavior and Associated Risk Factors. *Open Access Journal*, Vol 8, 110974-110985.
- [7] Yuan, J., Kim, H. and House, L. A. (2016). Zero inflated ordered probit approach to modeling mushroom consumption in the US, Food and Resource Economics Department, University of Florida. *International Food and Agribusiness Management Review*: 20 (5), 655 – 672.
- [8] Yong-Woo, L. (2016). State dependence, unobserved heterogeneity, and health dynamics in Korea. *Hitotsubashi Journal of Economics*, 57 (2), 195-221.
- [9] Heckman, J. (1981c). *Heterogeneity and state dependence*. In *Studies in Labor Markets*, Rosen (ed). University of Chicago Press.
- [10] Contoyannis, P., Jones, A., & Rice, R. (2004). The dynamics of health in the British household panel study. *Journal of Applied Econometrics*, 19, 473-503.
- [11] Akay, A. (2009). Finite-sample comparison of alternative methods for estimating dynamic panel data models. *Journal of Applied Econometrics*, 27, 1189-1204.
- [12] Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, 20 (1), 39-54.
- [13] Rabe-Hesketh, S., and Skrondal, A. (2013). Avoiding biased versions of Wooldridge's simple solution to the initial conditions problem. *Economics Letters*, 120 (2), 346-349.
- [14] Raymond, W., Mohnen, P., Palm, F., and Loeff, S. S. V. D. (2007): "The Behavior of the Maximum Likelihood Estimator of Dynamic Panel Data Sample Selection Models", CESifo Working paper No. 1992.
- [15] Mulkay, B. (2015). Bivariate probit estimation for panel data: a Two-step Gauss Hermite Quadrature approach with an application to product and process innovations for France. Universit'e de Montpellier, Montpellier.
- [16] Christelis, D., and Galdeano, S. A. (2009). Smoking persistence across countries: An analysis using semi-parametric dynamic panel data models with selectivity. *Journal of Applied Econometrics*, 15, 334-360.
- [17] Lee, W. S., and Oguzoglu, U. (2007). Well-being and ill-being: A bivariate panel data analysis (Vol. 30). IZA Discussion Paper Series No. 3108.
- [18] Kano, S. (2008). Like husband, like wife: A bivariate dynamic probit analysis of spousal obesities. Osaka Prefecture University.
- [19] Efron, B., and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected fisher information. *Biometrika*, 65 (3), 457-487.
- [20] Wooldridge JM. 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA.
- [21] Lesaffre, E., and Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random-effects model: an example. *Applied Statistics*, 50, 325-335.