CLASSICAL AND BAYESIAN APPROACHES FOR THE ZERO-INFLATED DYNAMIC CATEGORICAL PANEL ORDERED PROBIT MODEL

JOHN KUNG'U WANJIRU

I84/37359/2016

A THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF THE DEGREE OF DOCTOR OF PHILOSOPHY (STATISTICS) IN THE SCHOOL OF PURE AND APPLIED SCIENCES OF KENYATTA UNIVERSITY

JANUARY, 2023

DECLARATION

This thesis is my original work and has not been presented for a degree in any other University or any other award.

Signature..................................... Date............................

John Kung'u Wanjiru

I84/37359/2016

Department of Mathematics and Actuarial Science

Kenyatta University

SUPERVISORS

We confirm that the work reported in this thesis was carried out by the student under our supervision.

Prof. Leo Odiwour Odongo

Signature..................................... Date............................

Department of Mathematics and Actuarial Science

Kenyatta University

Dr. Ananda Omutokoh Kube

Signature....................................... Date..............................

Department of Mathematics and Actuarial Science

Kenyatta University

DEDICATION

To my family, Elizabeth Wanjiku, Joseph Maina and Samuel Mbiu.

ACKNOWLEDGEMENTS

I thank the Almighty God whose unending blessings and mercies have enabled me to accomplish this study.

I take this opportunity to convey my appreciation towards my supervisors Prof. Leo O. Odongo and Dr. Ananda O. Kube of Kenyatta University, Department of Mathematics and Actuarial Science who provided their professional support, guidance and moral support in writing this thesis. Their positive critiques, invaluable comments, timely feedback, guidance and patience enabled me to improve the quality of this thesis. My appreciations also go to the Department of Mathematics and Actuarial Science, which provided me with a conducive environment to undertake this work.

My gratitude also goes to my beloved wife Elizabeth Wanjiku who is always standing with me through good and bad times, for your support and love in our marriage. You are always my constant source of love, happiness, and strength all the years.

Finally, I pay rich tribute to my friends and workmates at Longonot Township Secondary school and Laikipia University for providing all kind of support including guidance, encouragement, time and above all their prayers.

LIST OF ABBREVIATIONS

OP	Ordered Probit
AOP	Autoregressive Ordered Probit
ZIOPC	Correlated Zero Inflated Ordered Probit
DPOP	Dynamic Panel Ordered Probit
ZIDPOP	Zero Inflated Dynamic Panel Ordered Probit
ZIDPOPC	Correlated Zero Inflated Dynamic Panel Ordered Probit
ZIDPOPI	Independent Zero Inflated Dynamic Panel Ordered Probit
MCMC	Markov Chain Monte Carlo
AIC	Akaike Information Criteria
DIC	Deviance Information Criteria
GHQ	Gauss-Hermite Quadrature
APEs	Average Partial Effects
RMSE	Root Mean Squared Error
GHK	Geweke-Hajivassiliou-Keane
SCMH	Single Component Metropolis-Hastings
NLSY97	National longitudinal survey of youth 1997

DEFINATION OF TERMS

Autoregressive model It is a model that uses previous observations in time series as an input to a regression equation to forecast the current observation of the time series.

Categorical variable It is a finite variable that represents two or more groups.

Cross sectional model It is a model applied when observations are assembled across respondents at an exact time only.

Dynamic panel model It is a model applied when observations are assembled across individuals and over a span of time.

Latent variable model It is any statistical model that considers the presence of a latent variable that expresses the association among response variables.

Ordered Probit model It is a model that supports the statistical association between dependent variable with independent variables as in ordinary least square regression. In contrast to ordinary least square regression it recognizes the uneven difference among the discrete ordered values.

Ordinal value A discrete variable whose values are well arranged with respect to a particular property.

State dependence It is the dependence of current choice on previous choice.

Zero inflated model It is a distribution that permit zero responses that are greater than expected.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

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 substances such as drugs. 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 assumes that the cut points are known to address heaping in the data and therefore cannot be used when the cut points are unknown. The Simulated maximum likelihood was applied to evaluate the double integrals in the Semi-parametric zero inflated dynamic panel probit model. This procedure could be very time-consuming even with fast modern computer and imprecise even with the use of modern simulator like Halton simulators. The aim of this research was to develop the Zero inflated dynamic panel ordered probit models with independent and correlated error terms to address the above challenges. Interpretation of the coefficients in the proposed models were extra difficult than in the normal regression scheme because a shift in one of the variables in the equation is conditioned on other variables and their parameters. Average partial effects that gave the effects on the particular probabilities per unit change in the covariates was proposed to address the above challenge. The integrals were evaluated using Two step Gauss Hermite quadrature that is five times faster than the Simulated maximum likelihood. Since the solutions are not of closed form, maximum likelihood estimation based on Newton Raphson algorithm and Bayesian approach were used to estimate the parameters of the proposed models. Monte Carlo simulations were conducted to investigate the theoretical properties of the estimators of the developed models. Using National Longitudinal Survey of Youth (1997) dataset sponsored by the Bureau of labour Statistics of the U.S. Department of labour with zero inflation, the study investigated the determinants of smoking tobacco among the youths. The study found that the proposed models produced consistent estimators and their estimates were more accurate than the Dynamic panel ordered probit model estimates. The proposed models fitted the data better than dynamic panel ordered probit model in both classical and Bayesian approaches in the simulated data. 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 essential to control for participation decision and consumption levels at the initial period. The models showed a strong and significant positive state dependence in both participation decision and at various consumption levels. The unobserved individual effects accounted for 49.90% of the unexplained variation in decision to participate in smoking and 47.65% of the unexplained variation at all levels of consumption. The main causes of persistence in smoking decision were the state dependence, unobserved heterogeneity and race while the main causes of persistence at consumption level were state dependence, unobserved heterogeneity, gender and age. The study is significant to policy analyst in identifying the socioeconomic and demographic factors associated with drug abuse and providing useful information to facilitate well-targeted public health policies.

CHAPTER ONE

INTRODUCTION

A brief discussion on the background of the study, statement of the problem, main objective, specific objectives, significance and the literature review are presented in this chapter.

1.1 Background

Many studies involve collecting data that are either continuous or discrete on one or more than one outcome from a given subject. The discrete data may be either nominal or ordinal. Categorical data includes gender, marital status, age group, satisfaction levels etc. Nominal outcome consists of outcome that represents groups or categories without any given order e.g. gender (male or female) while ordinal outcome represents groups or ranked categories e.g. education level (Primary, Secondary, Diploma, Degree, Masters, PhD).

Ordered categorical observations are assembled in various fields such as international conflict studies, behavioral research, biomedical research etc. According to Kostecki-Dillion *et al.* (1999), responses based on severity of migraine can be classified as none, mild, moderate, severe and intense and assumes values 0, 1, 2, 3 and 4 respectively. According to Murphy *et al.* (2008), responses based on the clients consumption of cocaine can be classified as none users, monthly users, weekly users and daily users and assumes values 0, 1, 2, 3 and 4 respectively. Similarly in social and behavioral research, attitudes of the respondents towards life satisfaction by Hasegawa (2009) or satisfaction with healthcare services by Gallefoss and Bakke (2000) and Williamson *et al.* (1995) are evaluated on a Likert scale from very dissatisfied to very satisfied. Self-assessed health condition by Contoyannis *et al.* (2004) is 1 denoting very poor to 5 denoting excellent.

Latent variable models are common in areas such as behavioral studies .It is a probabilistic model that decodes concealed forms in the observed data or representation that considers the existence of a latent variable that describe the association among reported observations. These models include structural equation models, latent class models and item response theory models. Bartholomew (1983) categorized the models into three classes' namely latent trait analysis, latent profile analysis and latent class analysis. In latent class analysis and latent trait analysis models, reported observations (indicators) are binary or ordinal variables and their conditional distributions are considered either binomial or multinomial. In latent class models, a continuous underlying latent variable is replaced by a discrete variable with "classes" that represents classes of homogeneous individuals. They identify unobservable sub groups within a population.

Liu and Agresti (2005) proposed to model ordered data in such a way that each ordinal variable is standing for a categorized form of an underlying continuous random variable. The ordinal data are created by putting cut points on continuous variables. For instance, the rank of a learner in a certain subject is A if the scores lies in the interval 70-100, B if the scores lies in the interval 60-69, C if the scores lies in the interval 50-59 , D if the scores lies in the interval 40-49 and F if the scores lies in the interval 0-39. The difference between the levels is inconsequential. It is only considered that an individual who choses one level had abundant characteristics in that level than if he/she had chosen a lower level, but the magnitude is unfamiliar to us. This approach is useful when the exact measurement is impossible or inconvenient; for instance, a patient depicts his/her health condition as poor, normal or good. There is a tendency of analyzing ordinal data as if they were continuous or binary with two responses, that is, Yes or No. The ordinal values do not have metric properties. Average, standard deviation and covariance of ordinal values lacks meaning due to absence of origin or unit of measurements. According to Fielding and Yang (2005), this may either lead to lose of information and efficiency or results into biased estimates and in some cases, numerical convergence might fail.

Time series data involving ordered categorical data are often encountered in numerous studies. In addition, temporal correlation normally occurs between adjacent observations in data gathered over period. The observations obtained by collecting the behaviour of entities e.g. individuals, countries etc. at cross sectional and multiple time units are referred to as Panel data. According to Park (2012), Panel data (also known as longitudinal) is a dataset in which the behaviours of units such as countries, companies, individuals etc. are surveyed across time. Its main benefit is the ability to permit greater adaptability in modelling variations in behaviours across individuals. This is because of collecting recurring observations from the same individual over time. Hence, the impetus for using a panel dataset is its ability to control for unobserved individual effects. The benefits from panel data occur in association with extra assumptions, which enlist some form of stability in the time pattern of options and constraints on the number of outcomes per unit. The panel data possess a large number of observations that gives more informative data, less multicollinearity, more degree of freedom and a higher efficiency of estimates. It is also possible to separate between cohort, period and age effects. The longitudinal data improve the possibilities of evaluating the effects of policy interventions and it is possible to determine under which conditions the effects can be interpreted as casual effects. The availability of panel data allows us to estimate treatments effects consistently without assuming ignorability of treatments and without an instrumental variable, provided the treatment varies over time and is uncorrelated with time varying unobservables that affect the response. The limitations of panel data are the presence of attrition, time varying sample sizes and structural changes. Panel surveys are also labour and cost intensive. The merits of panel model are now well grounded in empirical research such as household-level demand and labour supply decisions, workers' wage processes, firm-level productivity, or cross-country determinants of economic growth etc.

However, the empirical success of panel data is limited to linear models and special nonlinearities, for which a more or less complete understanding of identification and inference is available. Unobserved individual heterogeneity can be specified using fixed and random effects models. The fixed effects model permits the individual effects to be related with the covariates while random assume the individual effects are unrelated with the covariates. The limitation of a random effects approach is imposition of strong assumption of independence between the unobserved individual effects and explanatory variables. The initial condition problem is the other weakness of a random effects approach when estimating the dynamic ordered probit model. The random effect represents the random intercept in the model. However, the strong assumption in the random effects approach is relaxed in fixed effects approach. Despite this advantage, the fixed effect is rarely used due to the difficulty of solving the incidental parameters problem, that is, for every respondent that join the sample the number of parameters to be evaluated increases at a one-to-one rate as discussed by Neyman and Scott (1948). There is certainly no possibility of constructing a log-likelihood with a fixed T that allow us to consistently evaluate unobserved individual effect. It can only be constructed when T tend towards infinity. This implies that individual-specific fixed effects cannot be regarded as parameters to be evaluated due to incidental parameters problem. This influences the analysis of the degree of state dependence and average partial effects.

The ordered probit model and ordered logit model are commonly used to analyse the ordered categorical data. The logit link function is used by ordered logit model while the inverse normal link function is used by ordered probit model. Zero inflated Ordered probit model easily allows the modeling of the correlation between the error terms in binary probit model and the ordered probit model while Zero inflated ordered logit model cannot lend itself easily to allow for correlation between the error terms in binary probit model and the ordered probit model. The logit model cannot be used with panel data when unobserved factors are correlated over time.

The Ordered probit model has ability to provide the statistical significant relationship between the study variable with covariates as in ordinary least square regression. In addition, unlike ordinary least square regression, it is able to identify the unequal difference between the levels of the categories. OP model is stipulated in relation to correlated or uncorrelated univariate and multivariate normal distribution of the underlying continuous latent variables that appears as ordinal variables by discretization.

Often, when assessing unusual tendencies like consumption of prohibited drug, international conflict, severity of migraine etc. presence of zero inflation occur due to non-participation, presence of peace, absence of migraine, zero consumption etc. According to Lambert (1992), the zero inflation implies zero responses that are greater than expected or the number of zeros surpass what the model would normally produce. The zero inflated model is a statistical model that permit zero responses that are greater than expected and account for them by incorporating their origin.

Harris and Zhao (2007) showed that OP model has restricted ability in expounding the source of excessive zero responses, particularly when the zeros are related to the dual and separate systems. In the circumstances, where we are dealing with an unusual but legal behaviours such as smoking of cigarette, a zero may be recorded for non-user who reports zero consumption because of their religious norms, state of health or moral stand. Another zero may be recorded for an active user whose current consumption is zero due to government policy that ban the smoking of cigarette in certain places, low income, high cost of cigarette or in rehabilitation center. However, this user may become an active consumer if he/she is around the smoking zones, prices are lower or income is higher. Hence, the zeros from non-participant and participant who report zero consumption are influenced by distinct systems of consumer behaviours. The users who report zero consumption are possible users and possess characters identical to those of the active users and are likely to become active users when the price are lower or income is higher. The bona fide non-consumers are expected to possess an absolute inelastic cost and income demand schedules, and are influenced by distinct systems associated with norms, state of health and moral stands. If such underlying systems are not considered, it could nullify any ensuing policy issues. Similarly, the covariates, in addition, could exhibit distinct impacts on the binary underlying systems. For instance, the impact of income on smoking of cigarette, higher income, as a measure of wealth, may increase the likelihood of users who report zero consumption to smoke. However, for non-users, higher income will have an inelastic effect on them.

The traditional OP model that uses a single latent equation will conceal the differential effect between the non-user and active user whose current consumption is zero. Harris and Zhao (2007) proposed using the likelihood approach a ZIOPC model that considered the non-participation, zero consumption and active participation distinct systems of consumer behaviours and applied the binary probit model for participation level and OP model for consumption levels.

The Bayesian and Maximum likelihood estimations techniques are normally employed to evaluate the parameters of a model. In Maximum likelihood estimation, the parameters are fixed but unknown. We obtain point estimates while interval estimates relies on the accuracy of the procedure and not on the parameter. In the Bayesian approach, the parameters are random but still unknown. Here, we obtain the whole distribution of parameter estimates and their variability. The Bayesian statistics views probability as a degree of belief while classical views probability as relative frequency observed during many trials. Classical approach is interested in a certain interval's inclusion of a parameter's real value called confidence interval. Bayesian approach is interested in the probability of a certain interval's inclusion of a posterior distribution's mean called credible interval.

Dynamic latent variable models requires solving integrals that are intractable. Numerical approximations such as simulation and two-step Gauss-Hermite quadrature technique by Lee and Oguzoglu (2007) and Kano (2008) and Raymond *et al.* (2007) respectively represent possible solutions to this problem. In simulation method, as pointed by Mulkay (2015), the computation is very time-consuming and imprecise even with the use of modern simulator like GHK or Halton simulators. Gaussian quadrature proposed by Raymond *et al.* (2007) is favoured in evaluating the two-dimensional indefinite integral since it perform well for even a small number of nodes and weights.

According to Park (2012), Bayesian theory relies on prior distribution and likelihood function. The prior distribution relies on data collected from previous studies, expert opinion, distributions that denote preceding information or even a latent variable rather than an observable variable. It is then improved by merging it with observed data that is denoted by a likelihood function. This produces a posterior probability distribution that merges the prior distribution and likelihood function to form a single probability distribution. Prior and posterior distributions are called conjugate distribution in case they come from the same family such as normal distribution. The conjugate distribution produced have a closed form. There are two types of conjugate prior namely informative and non-informative conjugate prior. Informative conjugate prior convey specific and definite information about a variable. Non-informative conjugate priors are used when one has little or no knowledge about the data and hence has the least effect on outcomes of the analysis. Non-informative conjugate priors simplify resampling using MCMC algorithm and possess attractive convergence properties.

Albert and Chib (1993) introduced a data augmentation that simplify the fitting of the Bayesian probit model by simulating draws from the posterior distribution via Gibbs sampling. Bayesian inference heavily depends on high-dimensional integration over the posterior distribution to make inference on the parameters that make it difficult to calculate analytically. This difficult is solved by simulation-based integration called MCMC algorithm that permit us to draw samples from posterior distributions by creating a Markov chain that has the target distribution as its equilibrium distribution. The discovery of MCMC algorithms has made Bayesian approach more approachable due to availability of cheap and high computing power. The advantages of Bayesian approach are the ability to incorporate prior knowledge formally into data analysis and MCMC algorithms can easily obtain important parameters for policy decisions such as elasticity and marginal effects.

Gurmu and Dagne (2012, 2009) extended Harris and Zhao (2007) work to Bayesian inference for univariate ZIOPC model and Bivariate ZIOPC model respectively. They demonstrated using the Deviance Information Criterion that accounting for the zero inflation in the ZIOPC present a better fit than ordered probit model. They also found that ignoring the zero inflation may lead to model misspecification and concealing the differential results independent variables have on non-participant versus participant at different stages of consumption, including non-consumption. In addition, if the two zeros are not modelled correctly, it could invalidate any subsequent policy implications.

There are two types of persistence revealed by Heckman (1981a) in labor market. The first type of persistence is due to observed heterogeneity such as level of education, age, marital status, health problems etc. and unobserved heterogeneity such as self-drive, social capital etc. These observed heterogeneity and unobserved heterogeneity may induce recurring unemployment across periods. If left unaccounted for, the heterogeneities induce false state dependence in labor market histories. In the second type of persistence, previous unemployment period may itself have an effect on the chance of current employment. The second type of persistence is called the true state dependence.

Contoyannis *et al.* (2004), Ayllon and Blanco (2012) and Yong-Woo (2016) used DPOP model to study the dynamic self-assessed health in Britain, Spain and Korea respectively. As indicated by Ayllon and Blanco (2012), the effects of observed heterogeneity on self-assessed health are generally overrated since they detect the impact that should be associated with past health or other unseen heterogeneities absent in majority of datasets. Their results indicated that state dependence and unobserved heterogeneities accounted for much of the chances of observing a specific health status while the importance of observed heterogeneities diminishes when controlling the state dependence and unobserved heterogeneities. The unobserved heterogeneity accounts for 30%-40% of reporting a specific health status. Heckman (1981b) pointed out that poor handling of unobserved heterogeneities produces a conditional association between previous and current experiences, which is referred to as false state dependence. The separation of true and false state dependence is vital in policy-making.

Heckman (1981b) and Wooldridge (2005) proposed two ways of treating the initial conditions problem. Heckman (1981b) used reduced-form equation that rely on available pre-sample information as an approximation method for the conditional distribution of the first observation. The drawback of this approach was high computational cost in the estimation process as approximation of the conditional probability of initial values leads to a simultaneous estimation problem of the reduced-form and structural model. Wooldridge (2005) approach was based on unobserved individual-effects conditional on the initial values and time variant variables. As the number of time variant variables and/or T grows, the model will possess a huge number of variables that will decrease the degrees of freedom and make it difficult to evaluate the integrals. This will intensify evaluation time significantly, even for relatively moderate panels. This approach results into a tractable functional form and consistent estimator. The two approaches produces consistent estimates of the parameters of the model under the assumption of correct specification of the distribution of the errors. However, according to Raymond *et al.* (2007), the Wooldridge (2005) approach is easier to apply and more flexible in the sense that it handle a vast range of nonlinear dynamic panel data models and permits, unlike the Heckman approach, for individual effects to be associated with the strictly exogenous covariates. The likelihood function for dynamic and the static versions within Wooldridge approach have a similar structure in nonlinear model.

Akay (2009) proposed a constrained model where the exogenous variables in the Wooldridge (2005) methods are replaced by the within-means based on all periods including the first, and thus reduces the number of included variables. Wooldridge (2005) indicated that consistency requires correct specification of the conditional distribution of the unobserved effect given initial observation and exogenous variables. Rabe-Hesketh and Skrondal (2013) revealed that the constrained model proposed by Akay (2009) could be severely biased because it absolutely fixes 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 motive is that the conditional distribution of the unobserved effect, given the covariates at all periods (including the initial period), relies more directly on the initial-period covariates than on the covariates at the other periods — in some cases it relies only on the initial-period covariates and the initial dependent variable. The coefficients of the initial-period covariates should therefore not be constrained to be equal 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.

Gurmu and Dagne (2009) and Harris and Zhao (2007) ZIOPC models ignored the unobserved heterogeneity and true state dependence that account for much of the chances of observing a specific consumption level of cigarettes. Harris and Zhao (2007) analysed smoking data using ZIOPC models from Australian National Drug Strategy Household Survey. The study used surveys of 1995, 1998 and 2001. Although the smoking data was panel, Harris and Zhao (2007) ignored state dependence, unobserved heterogeneity and initial condition problems and hence overestimated the effects of the observed heterogeneity.

Christelis and Galdeano (2009) proposed a Semi-parametric zero inflated dynamic panel ordered probit model with selectivity to analyse smoking persistence across countries. They used Akay (2009) approach in modeling the relationship between the initial observation and unobserved individual effect. They assumed that the individual effect follows a K-point nonparametric distribution. Their cut points are known because they are chosen to cater for bunching for cigarettes smoked The knowledge of cut points was substantial. The OP model estimates the ratio between cut points and standard deviation and the knowledge of cut points allowed them to identify the standard deviation.

The constrained model proposed by Akay (2009) and used by Christelis and Galdeano (2009) to modeling the initial condition was shown by Rabe-Hesketh and Skrondal (2013) to be severely biased because it absolutely fixes the coefficients of the initial explanatory variables to be equal to the coefficients for the subsequent periods, which is at odds with the form of the correct distribution. This indicate that the Semi-parametric Zero inflated dynamic panel ordered probit model with selectivity proposed by Christelis and Galdeano (2009) was severely biased. The model assumes presence of heaping and hence the cut points can be deduced to handle the heaping. This model cannot be used when the cut points are unknown. The non-parametric distribution of the individual effects does not include the association between the unobserved individual effects and their variances. If the correlation is significant between the unobserved individual effects, then the unobserved individual effects factors affecting participation decision have an effect on an unobserved individual effects factors affecting consumption as well. Disregarding this association can result to inconsistent estimates. The variances of the individual effects allow the estimation of the inter-unit correlation coefficient that determine the percentage of latent error variance associated with unobserved individual effects.

This study proposed parametric ZIDPOPC and ZIDPOPI models with unknown cut points that examined how state dependence, unobserved heterogeneity, observed heterogeneity, initial condition using Rabe-Hesketh and Skrondal (2013) approach and zero inflation jointly affect overall evolution of time series and applied them to model cigarettes consumption among youths. The cut points in ZIDPOPC and ZIDPOPI models represents the position on the latent scale where a respondent shifts from one category to another.

1.2 Statement of the Problem.

The Semi-parametric Zero inflated dynamic panel probit model with selectivity was based on a constrained approach for initial conditions that lead to biased estimates because it absolutely fixes 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, non-parametric distribution for unobserved individual effects and known cut points. The model cannot be applied in data with unknown cut points. The model ignored the correlation between the time invariant heterogeneity terms in the participation and consumption equations and therefore assumed that they were uncorrelated. This assumption is not trivial. For example, if this correlation is significant, then it implies that factors affecting participation have an effect on cigarette consumption as well. Ignoring this effect can lead to inconsistent estimates and overestimation of the estimated standard deviations of the individual effects. The model ignored the variances of unobserved heterogeneities that are used to determine the latent error variance associated with individual effects. The model used simulated maximum likelihood approach where the individual's effects are integrated out by computing the double integral by simulation. This procedure could be very time-consuming even with fast modern computer and imprecise even with the use of modern simulator like GHK or Halton simulators. Simulated maximum likelihood require computation of a large number of cumulative density function in order to obtain sufficient precision in the log-likelihood function. The simulated probabilities are different each time it is resimulated with a new set of random draws producing a simulation noise. This simulation noise decline, for a given sample size, as the number of draws increases, becoming trivial if the number of draws is large enough.

This study proposed Zero inflated dynamic panel ordered probit model with independent and correlated parametric errors terms that incorporates the state dependence, unobserved heterogeneities based on parametric distribution with selectivity, unconstrained initial conditions and unknown cut points. This thesis used an alternative approach based on a two-step Gauss-Hermite Quadrature to evaluate the two integrals and is five times faster than the simulated maximum likelihood. The two-step Gauss-Hermite Quadrature relies on a decomposition of the two-dimensional normal distribution for the individual effects into a one-dimensional marginal distribution and a one-dimensional conditional distribution. The two-step Gauss-Hermite Quadrature have satisfactory performance in finite sample for even a small number of nodes and weights. Due to intractability of the model, parameters were estimated by maximum likelihood estimation based on Newton Raphson algorithm. However, the classical approach involve huge computation cost, sometime numerical convergence might fail and does not allow incorporation of prior information. Bayesian approach was proposed to address the computation and convergence issues in classical approach since it allow sampling from the conditional distributions and does not require the computation of first and second order derivatives and incorporate of prior information.

1.3 Objectives.

1.3.1 Main Objective

The aim of this research was to develop Zero inflated dynamic panel ordered probit model with independent and correlated error terms, their estimation based on maximum likelihood and Bayesian approaches and their applications in smoking data.

1.3.2 Specific Objectives

i To develop and investigate the properties of Zero inflated dynamic panel ordered probit model with independent and correlated error terms.

ii To develop a maximum likelihood estimation for Zero inflated dynamic panel ordered probit model with independent and correlated error terms.

iii To develop a Bayesian estimation technique for Zero inflated dynamic panel ordered probit model with independent and correlated error terms.

iv To evaluate the performance of the Zero inflated dynamic panel ordered probit model with independent and correlated error terms against the Dynamic panel ordered probit model based on simulation study and on an empirical data.

1.4 Significance of the Study

The investigation used Rabe-Hesketh and Skrondal (2013) approach for initial conditions that reduced the bias introduced in the study by a constrained model proposed by Akay (2009). The study used a bivariate normal distribution that is parametric for unobserved individual effects. The parametric distribution facilitate the estimation of variances of unobserved heterogeneities that are used to determine the percentage of latent error variance related to unobserved heterogeneities and the correlation coefficient between the unobserved individual effects determine whether issues influencing participation decision have an impact on consumption levels as well. The study facilitate the estimation of the unknown cut points. The two integrals are evaluated using two-step Gauss-Hermite Quadrature that have satisfactory performance in a finite sample for even a small number of nodes and weights and is five times faster than the simulated maximum likelihood. The Gauss-Hermite Quadrature sometime may result in high computational cost. The study has presented a Bayesian approach that reduces the high computational cost by sampling from conditional distributions. The study is significant to policy analyst in identifying the socioeconomic and demographic factors associated with drug abuse and providing invaluable information to facilitate well-targeted public health policies.

CHAPTER TWO

LITERATURE REVIEW

2.1 Classical Approach

Olsson (1979) and Muthen and Kaplan (1985) discussed the treatment of ordered categorical data as continuous through analytical and simulation respectively. They found that a small bias was introduced in the presence of at least five categories and when the observations have a small symmetric distribution. A smaller bias was introduced in the presence of at least seven categories. However, in the presence of at most four categories or a skewed distribution, a downward bias was introduced in the parameters and their standard errors. Such cases demand a special statistical method for the ordinal data. A well-designed approach assume that the reported ordinal data are linked to continuous latent variables. This was induced by use of cut points that partition the continuous latent variable into serialized regions matching every ordinal category.

Bliss (1934) first proposed the probit model for binary data. It changed the sigmoid dose-response curve to a linear graph, which could later be analyzed by regression estimation. The key attribute in probit models was the assumption of a latent variable that determines the level of the observed response through thresholds. The utility of the model is not affected when the existence of the latent variable does not seem natural. The probit model has recently been used by Kitenge (2020) to investigate household income patterns of COVID-19 infections. He found that the chance of acquiring COVID-19 infection does not rely on whether an individual is rich or not. Swinerton (2021) used the probit model to analysis how employment was affected by COVID-19 infections. The study revealed that towards the end of 2020, the ability to work from home or working for an essential business such as health reduced the chances of unemployment.

Aitchison and Silvey (1957) proposed a probit model for ordinal data to test the feedback of separate doses of stimulus that were categorized into ordinarily ranked group. The extension of probit model to accommodate multiple independent variables was done by McKelvey and Zavoina (1975) and they also presented the original detailed discussion of ordered feedback model. McCullagh (1980) utilized latent variable model that regarded the ordered feedbacks as grouped data from an underlying latent variable with the cut points for the groupings seen as unknown coefficient.

Martin and Concepcion (2021) used OP model to explain the citizens' satisfaction during COVID-19 pandemic by attitude and a number of socio-demographic factors within European Union countries. The study revealed that Danes, Irelanders, Greeks, and Croats were the most satisfied citizens in comparison with Spaniards who were the least satisfied nationals. The social class, political support to the government and level of education were significant determinant of satisfaction levels. Furthermore, most criticism came from respondents who were concerned by the state of economy and protection of individual rights than respondents who were concerned by the state of health. Das *et al.* (2021) used OP model based on telephone survey to investigate students' assessment for online class in the course of Covid-19 pandemic. The survey revealed that sex, cost of online classes, internet network, living area, skills of the tutors, academic grade and contemplation of online class as a better choice significantly affected the students' perception of online class.

Sometime, the count and ordinal data observations may contain excess zeros. Zero-inflated or hurdle count data models are commonly used to analyse count data in presence of excess zeros. Lambert (1992) developed Poisson Regression model with excessive zeros and applied it in quality control to determine the defects in Manufacturing. The zero inflation in count data can also be handled by hurdle count model proposed by Cragg (1971) as an example of truncated models that permits stochastic processes for various levels of participation.

In case of zero inflation in the ordered discrete choice models, Harris and Zhao (2007) developed a ZIOP model that account for the source of excess zeros using the binary probit distribution in a way identical to Lambert (1992) and Cragg (1971) approaches.

Unlike Lambert (1992) and Cragg (1971) approaches that have the Poisson regression framework, the ZIOP model comprises of a probit "splitting" model, which produced the non-participants and participants, and an OP model, which produced zero-consumption and non-zeros consumption participants. Furthermore, Harris and Zhao (2007) also defined the ZIOPC model that permit the error terms of the binary probit model and OP model to be correlated.

They demonstrated the merits of the ZIOPC model in isolating the diverse behavioral schemes for non-smokers and smokers. Specifically they permitted the splitting of zero users into non-smokers who choose not to smoke due to non-economic factors such as morals and smokers with zero consumption that may be the result of a demand-schedule corner solution and are therefore responsive to economic factors such as income, prices etc. They demonstrated that the use of OP model would confuse the effects of some key covariates that have contrasting results on the two schemes. The model can also be used to determine the proportion of zeros coming from each regime, and how this split shifts with observed characteristics. The ZIOP(C) model permits for the recognition of covariates that are vital in each regime for key policy analysis.

Several researches have been conducted using the ZIOPC model. They include Downward *et al.* (2011) in modelling sports participation, Bagozzi *et al.* (2015) in Modelling Two Types of Peace in international Conflict Research, Yuan *et al.* (2016) in modeling mushroom consumption in the USA and Acero and Luis (2019) modelling persistence in the imitation of innovations in products in the manufacturing industry of Colombia. Greene *et al.* (2018) have integrated misreporting in ZIOP model where the zero responses are permitted to originate from the nonparticipants, participant misreports' (who have larger loss functions associated with a truthful response) and irregular clients and applied it to cannabis consumption. The study revealed that the incidences of cannabis use was significantly affected by misreporting. The study also revealed that the degree of misreporting was affected by the administration of the survey, those in attendance when the survey was done and the respondent's general trust in the surveys.

The OP model deals with cross sectional dataset and therefore cannot adequately handle dynamic dataset that is collected over a period. In order to overcome this inadequacy, Jackman (2000a, 2000b) developed a dynamic probit model for ordinal data as an answer to a problem with quantitative research of international conflict and In and out of war and peace with transitional models of international conflict using binary probit model. He found that conditioning on the previous status was a key guiding principle in the study of dynamic data. Beck *et al.* (2012) proposed another model of dynamic binary model and applied it to "state failure," which seizes severe political crisis represented by such recent events as Afghanistan, Somalia and Bosnia. This weakened the establishments in charge of governance to an extent that they no longer exercise civil authority or uphold political order. They showed that overlooking dynamism in ordinary probit model may results into misleading inferences. A model that includes a lagged latent dependent variable and transition models provides sensible results. They claimed that the utilization of a lagged latent variable is often better than the use of a lagged realized dependent variable.

The AOP model has also been used to model rainfall data by Varin and Vidoni (2006) and to model migraine severity by Varin and Czado (2010) using pairwise likelihood function. Brooks *et al.* (2008) proposed an inflated OP model for monetary policy, that is, a binary probit model that indicates the inclination of monetary policy to change or not and an OP model showing the direction of change based on Taylor (1993) type variables. They inflated OP model did not include autocorrelation that is likely to be present between adjacent observations. They also accounted for unobserved heterogeneity that exists between multiple responses. David and Sirchenko (2018) developed a model for ordinal feedbacks to determine the policy interest rate. The heterogeneity was captured with two-stage cross-nested model that combined three ordered probit equations. The model permitted the existence of the possible correlation amongst the three latent decisions. Monte Carlo simulations demonstrated suitable outcome in small samples. The results based on panel data on individual policymakers' votes for the interest rate showed superiority with respect to the conventional two-part OP model.

Christelis and Galdeano (2009) proposed a Semi-parametric ZIDPOP model that was an extension of ZIOPC model by Harris and Zhao (2007) within a semi-parametric dynamic framework. It permitted the time-varying error terms between participation decision and various consumption levels to be correlated. They accounted for unobserved heterogeneities in a semi-parametric approach for the sake of differentiating genuine state dependence from non-genuine state dependence. The model demonstrated that the presence of unobserved individual heterogeneity results into a huge reduction in the size of the genuine state dependence.

Contoyannis *et al.* (2004), Ayllon and Blanco (2012) and Yong-Woo (2016) used DPOP model to study the dynamic self-assessed health in Britain, Spain and Korea respectively. They found that unobserved heterogeneity and state dependence explain much of the chances of registering a particular health status while the importance of observed heterogeneity diminishes when managing both. Overlooking the state dependence overestimate the effect of observed heterogeneity while ignoring unobserved heterogeneity will overestimate the state dependence. According to Heckman (1981a), the poor treatment of unobserved individual effects gives rise to a conditional association between previous and current experiences, which is referred to as false state dependence. To differentiate the true state dependence from the false state dependence is of great concern to policy-making. The presence of state dependent variable as covariates and absence of the first observation in the panel data results in the initial conditions problem. This refers to the challenge of correctly specifying our observed first period distribution when data is not observed from time 0. This results in biased estimation in short dynamic panels. Heckman (1981b) and Wooldridge (2005) approaches are the commonly used methods of dealing with the initial condition problem. Heckman (1981b) approach specifies a model for the initial conditions given the individual effects and the strictly exogenous covariates. The model is often taken to be similar to the model underlying the remaining process. This results into huge computational cost. Wooldridge (2005) proposed to model the distribution of the unobserved heterogeneity conditional on the first reported state and time-varying covariates at each period. As the number of time variant variables and/or T grows, the model will contain a huge number of variables that will lower the degrees of freedom, make it difficult to evaluate the ensuing integrals and escalate estimation time for even a moderate panel.

Akay (2009) proposed a constrained model where the exogenous variables in the Wooldridge's methods are replaced by the within-means based on all periods including the first, and thus reduces the number of included variables. Wooldridge (2005) revealed that consistency demands correct specification of the conditional distribution of the unobserved heterogeneity given first observation and exogenous variables. Akay (2009) found that Wooldridge (2005) approach for initial condition problem perform well for the panels with at least 5 periods while Heckman's reduced-form approximation perform well for the panels with less than 5 periods. This is because the Wooldridge method does not specify an explicit conditional probability distribution for the first values, and the bias obtained with this method is behaviourally the same as the exogenous first values assumption for very short panels. He showed that all the approaches works equally well for panels with at least 10 periods.

Rabe-Hesketh and Skrondal (2013) revealed that the constrained model proposed by Akay (2009) could be severely biased because it absolutely fixed the coefficients of the first covariate equal to the coefficients for the succeeding periods, which is at odds with the form of the correct distribution. The rationale is that the conditional distribution of the unobserved heterogeneity, given the covariates at all periods (including the first period), relies more directly on the first-period covariates than on the covariates at the other periods — in some cases it relies only on the first-period covariates and the initial dependent variable. The coefficients of the first-period covariates should therefore not be constrained to equal the coefficients at the other periods. They showed that the bias for the constrained model practically vanishes when the initial period covariates are added as additional covariates or by using Wooldridge's original approach.

2.2 Bayesian Approach

Albert and Chib (1993) developed the Bayesian approach for the estimation of the OP model by utilizing the MCMC method and Gibbs sampler while operating the data augmentation method of Tanner and Wong (1987). They used the latent variable representation for the estimation of OP model. The shortcoming of their data augmentation was the high autocorrelation in the estimated cut points, that is, suffers from slow mixing due to a high correlation between the simulated cut off points and latent variables.

Cowles (1996) revealed that the sampling of the cut points conditioned on the latent data could result to minor shifts in the cut points between successive iterations particularly when more data become availed. The high autocorrelation could then also affect the convergence of regression coefficients. Cowles (1996) proposed to solve the problem by sampling the latent variable and the cut points jointly, that is, sampling the cut points given the observed ordinal value and the other parameters, marginalized over the latent variable and subsequently sampling the latent variable given the remaining parameters including the cut points and data. Since the resulting distribution of the cut points is not of standard form, he employed a sequence of Metropolis Hasting steps to sample each cut point conditioned on the remaining parameters, cut points and the data. Chib and Albert (2001) simplified the sampling of the cut points by transforming them so as to get rid of the ordering constraint by one-to-one mapping.

Butller (2011) proposed hierarchical probit models for binary and ordered ratings data to assess instructor effectiveness based on Bayesian approach. In order to account for the differences in instructor and course, they modified the latent variable structure by including an instructor by course term as well as the student confounders. By modifying the binary response to an ordered response, they were able to differentiate poor, average, and exceptional instructors. These models provide more insight into instructor effectiveness than the average scores alone and would be an effective tool for making recruitment and promotion decisions. Ohh (2015) also used the OP model using Bayesian inference to compare the three diagnosing tool for prediction of metabolic-related diseases. The study revealed that Waist-To-Height Ratio was significantly better than Body Mass Index and Waist Circumference for prediction of metabolic-related diseases in people aged above forty in Korea. Caglayan and Van (2017) determined the variables influencing the economic development levels of particular countries using Bayesian ordered probit model. They found that internet users, share of expected years of schooling, gross domestic product, health expenditure, seats in parliament had a positive effect, but life expectancy at birth and rural population had a negative impact on human development in long term.

Gurmu and Dagne (2012, 2009) extended Harris and Zhao (2007) work to Bayesian inference for univariate and Bivariate zero-inflated ordered probit and applied the model in modelling tobacco consumption. They found a powerful proof that accounting for zeros inflation presented a better fit to the data. They also found evidence that the use of a model that ignores excess zeros hides differential effects explanatory variables have on the two regimes, non-participant versus participant at various level of consumption, including zeros.

Several researches have been conducted using the ZIOPC model based on Bayesian inference. Oh *et al.* (2012) modelling Korean alcohol consumption. They found that the marginal effect of each covariate shows that certain covariate have effects on the genuine non-participant and potential participants in opposite directions, which may not be discovered by an OP model. Several authors have extended the ZIOPC Model to bivariate case based on Bayesian inference, these include Gurmu and Dagne (2012) and Rajendra (2013) who proposed the zero-inflated bivariate ordinal data using latent mixture approach that can be used in circumstances where two related feedback such as marijuana and cocaine are gathered simultaneously.

The OP model based on Bayesian inference cannot adequately handle dynamic data set that is collected over a period. In order to overcome this inadequacy, Muller and Czado (2005) proposed an AOP model based on Bayesian approach using ordinal data to cater for the autocorrelation between adjacent observations. They applied the model in analyses of high frequency financial data. Using the Bayes factor, they showed that the AOP model provided a better fit than an ordinary OP model.

Hasegawa (2009) proposed the DPOP model based on Bayesian approach and applied it to determining subjective well-being. His ordinal responses were from strongly dissatisfied to strongly satisfied. The study revealed that savings and income had positive effects while labor force participation and marriage had negative effects on life satisfaction in terms of the response probabilities.

Park (2012) developed the Bayesian inference of time-series and ordered panel data models with non-ignorable missing data and applied it in determining the determinants of self-rated health based on the Health and Retirement data. Stegmueller (2013) proposed Bayesian robust dynamic latent ordered probit model for modeling dynamic preferences and applied it on individuals' preferences for government intervention over a period of nineteen years. His results demonstrated the need of using a hierarchical dynamic panel modeling approach.

Seibert *et al.* (2018) developed the Cross-nested Autoregressive Ordered Probit (CronAOP) model to model policy interest rates. They opted for a Bayesian approach utilising MCMC methods and a Gibbs sampler with data augmentation, which made the estimation to be computationally feasible and without high-dimensional integration or numerical optimization. The simulations demonstrated that the developed Cross-nested Autoregressive Ordered Probit (CronAOP) model works well in small samples.

Maximum likelihood estimation of dynamic latent variable models require evaluation of integrals that are intractable. Numerical approximations provide a possible answer to this challenge. Lee and Oguzoglu (2007) and Kano (2008) proposed a simulated maximum likelihood approach for multivariate probit model where the individual's effects are integrated out by computing the double integral by simulation. Mulkay (2015) has shown that this procedure could be very time-consuming even with fast modern computer. In this thesis, an alternative approach based on a two-step

Gauss-Hermite Quadrature was used in order to evaluate this double integrals. The Gauss-Hermite quadrature was proposed by Butler and Moffitt (1982) to evaluate multi-dimensional indefinite integrals. The Raymond *et al.* (2007) used two-step Gaussian Hermite quadrature to implement maximum likelihood estimation of the dynamic panel data Type 2 and 3 Tobit models. A Monte Carlo study showed that the quadrature have satisfactory performance in finite sample for even a small number of nodes and weights. According to Raymond *et al.* (2007), accounting for individual heterogeneities but using exogenous initial conditions also leads to false state dependence situation. The state dependence can only be estimated if we take into account the association between the initial conditions and the individual effects. Mullahy (1997) proposed the bivariate probit estimation for panel data using the same approach and applied it to product and process innovations in France. A simulation shows the importance of estimating the correlation in random effects and the correlation between both equations.

The two-step Gauss-Hermite Quadrature algorithm results into high computational cost. This cost of direct computing the multiple integrals can be reduced in a Bayesian approach. Bayesian analysis is commonly used in many empirical studies due to some advantages that it has over the classical methods. These includes; the use of prior information that can generate more accurate results, it yields equivalent results to maximum likelihood methods under non informative priors, it does not depend on large sample theory and it can lower the computational cost in certain cases especially the one involving multi-integrals.

Harris and Zhao (2007) dataset from the three most recent surveys of 1995, 1998 and 2001 had a time structure. The ZIOPC model proposed by Harris and Zhao (2007) only accounted for observed heterogeneity but ignored the state dependence and observed individual effects that account for much of the chance of observing a specific status. This implied that they overestimated the effect of observed heterogeneity. They also failed to consider initial condition that is likely to produce inconsistent estimates.

Christelis and Galdeano (2009) proposed Semi-parametric zero inflated dynamic panel probit model with selectivity and applied to study smoking persistence across countries. They found a smaller true state dependence in participation decision and consumption levels when unobserved individual heterogeneity was included and they also uncovered huge disparities in true state dependence across countries. The fitting was considerably improved by considering the bunching of the reported amount of smoked cigarettes. The correlation between the participation and consumption levels was significant, this implied that factors affecting participation decision had an effect on consumption levels as well. Ignoring this association can results into inconsistent estimates. They used Akay (2009) approach in modeling the relationship between the initial observation and unobserved individual effect. They assumed that the individual effect follows a K-point nonparametric distribution. Their cut points are known because they are chosen to tackle bunching in the amount of smoked cigarettes. The knowledge of the cut points has a significance practical implications. Given that the ordered probit estimates the ratio between cut points and standard deviation, they knowledge of cut points allowed them to identify standard deviation of the dynamic panel ordered probit model.

Rabe-Hesketh and Skrondal (2013) revealed that the constrained approach could be severely biased because it absolutely fixed 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. Christelis and Galdeano (2009) model assumed the cut point are known and thus cannot be used in case where the cut points are unknown. They non-parametric distribution of the unobserved individual heterogeneity ignored the estimation of standard deviations and correlation of the individual effects. These standard deviations allow the estimation of inter-unit correlation coefficients that is used to determine the latent error variance credited to individual effects. If this correlation between the individual effects is significant, then it indicates that factors influencing participation have an effect on consumption levels as well. Ignoring the correlation between the individual effects may lead to overestimation of the estimated standard deviations of the individual effects and hence produce inconsistent estimates.

The parameter were estimated by the simulated maximum likelihood estimation. This method has been shown to be very time-consuming even with fast modern computer. Simulated maximum likelihood require computation of a large number of cumulative density function in order to obtain sufficient precision in the log-likelihood function. The simulated probabilities are different each time it is resimulated with a new set of random draws producing a simulation noise. This simulation noise decline, for a given sample size, as the number of draws increases, becoming trivial if the number of draws is large enough. The two-step Gauss-Hermite Quadrature relies on a decomposition of the two-dimensional normal distribution for the individual effects into a one-dimensional marginal distribution and a one-dimensional conditional distribution. The application of Gauss-Hermite quadrature, as an alternative to simulated maximum likelihood in this thesis, is inspired by the results of Guilkey and Murphy (1993) that, for the identical accuracy, the Gauss-Hermite quadrature method is 5 times as fast as the simulated maximum likelihood. The classical approach involve huge computation cost and sometime numerical convergence might fail. It does not allow incorporation of prior information in the estimation. Bayesian approach was proposed to address the compuation and convergence issues in classical approach since it allow incorporation of prior information and sampling from the conditional distributions.

This study developed a Zero inflated dynamic panel ordered probit models with unknown cut points that incorporate the state dependence, unobserved heterogeneities that have parametric distributions and initial conditions based on Rabe-Hesketh and Skrondal (2013). The parameters were estimated using maximum likelihood estimation and Bayesian approach. This model was then used to identify the determinants of smoking among the youths using the National Longitudinal Survey of Youth 1997 (NLSY97).

CHAPTER THREE

STATISTICAL PRINCIPALS FOR ANALYZING BINARY AND ORDERED PROBIT MODELS

3.1 Introduction

The cross sectional and dynamic probit models and their estimation techniques such as Maximum likelihood and a Bayesian approach based on Metropolis Hasting and Gibb sampling are presented. The cross sectional studies involves collecting data from a sample at an exact point in time while longitudinal involves repeatedly collecting data from the same sample over an extended period of time.

3.2 Binary and Ordered Probit ModeIs

According to Xin-She (2020), Binary data occurs when the variable of interest y_i assumes merely two values, i.e. $y_i \in \{0,1\}$, where $i = 1,2,...,n$ relate to items in the sample such as persons, household, firms etc. In term of notations, $y_i = 1$ typically denotes the existence of the incident of concern, whereas the absence is denoted by $y_i = 0$.

Ordinal data occurs when the variable of interest y_i can take one of the K ordered values, i.e. $y_i = 1,2,...,K$. The unique characteristic of ordinal data is that even though the responses are monotone, the scale on which they are evaluated is not considered to be cardinal and variations between categories are not directly comparable. The ordinal values should be mutually exclusive and exhaustive. The setup encompasses utility maximizing decision makers, who choses among contesting choices related to certain levels of utility. Specifically, respondent i has two levels of utility U_{i1} and U_{i0} that are related to $y_i = 1$. The utility maximizing agent then handpicks the preference granting the higher of the two utilities:

$$y_i = \begin{cases} 1 & \text{if } U_{i1} > U_{i0}, \\ 0 & \text{if } otherwise. \end{cases} \tag{3.1}$$

The utility U_{ij} is familiar to the respondent but are unfamiliar to the analyst, who can merely monitor a vector x_i of characteristics of the respondent that can be associated to utility through $U_{ij} = x_i\beta_j + \varepsilon_{ij}$ for $j = 0, 1$. β_j represents regression coefficient. x_i represents the covariates. The error term ε_{ij} captures the unobserved features that influence utility but are not embraced in x_i. This setup will be utilised to make probabilistic statements concerning the reported alternatives y_i conditionally on x_i.

To develop a model for the reported selections, given the covariate x_i and the parameters β_0 and β_1, the conditional probability of observing $y_i = 1$ can be specified as an exceedance probability between the two utility levels.

$$\begin{aligned} P(y_i = 1|x_i, \beta_o, \beta_1) &= P\left(x'_i\beta_1 + \varepsilon_{i1} > x'_i\beta_0 + \varepsilon_{i0})\right) \\ &= P\left(\varepsilon_{i0} - \varepsilon_{i1} < x'_i(\beta_1 - \beta_0)\right) \end{aligned} \tag{3.2}$$

It is actualized by expressing a density for the random variable $\varepsilon_{i0} - \varepsilon_{i1}$. From equation (3.2), we note that the selection probability relies solely on the differences in utilities between choices, not on the absolute level of utilities. Specifically because the probability in (3.2) relies on the difference $\beta_1 - \beta_0$, it will remain constant if we include an arbitrary constant d to both β_0 and β_1, i.e, $x'_i(\beta_1 - \beta_0) = x'_i((\beta_1 + d) - (\beta_0 + d))$. In addition, the scale of utility is not recognized since the probability remains constant if both sides of (3.2) are multiplied by an arbitrary constant $d > 0$, i.e,
$P(\varepsilon_{i0} - \varepsilon_{i1} < x'_i\beta_1 - x'_i\beta_0) = P(d(\varepsilon_{i0} - \varepsilon_{i1}) < dx'_i(\beta_1 - \beta_0))$. This challenge is dealt with by securing the location and scale of the utility. The location is secured by evaluating utility relative to that of the baseline category, U_{i0}. Specifically, we operate with the differenced form $z_i = x'_i\beta + v_i$ where $z_i = U_{i1} - U_{i0}$, $\beta = \beta_1 - \beta_0$ and $v_i = \varepsilon_{i1} - \varepsilon_{i0}$.

The association between the reported response y_i and the latent variable z_i is expressed as

$$y_i = \begin{cases} 1 & \text{if } z_i > 0, \\ 0 & \text{if } otherwise. \end{cases} \tag{3.3}$$

The scale is normalized by securing the dispersion measure of v_i and assuming it as available rather than an estimated value. This normalization does not constrain the underlying flexibility of the model. The fixed variance of v_i is model specific.

According to Xin-She (2020), a probit model is assumed if the error has a standard normal distribution with probability density function and cumulative distribution function given by $\phi(v_i) = \frac{1}{2\pi}\mathrm{e}^{-\frac{v_i^2}{2}}$ and $\Phi(v_i) = \int_{-\infty}^{v_i}\phi(t)\,dt$ respectively. The probability density function $\phi(v_i)$ is symmetric and the variance of v_i is secured at one as a normalization. Therefore,

$$\begin{aligned} P(y_i = 1|\beta) &= P(z_i > 0) \\ &= P\left(x'_i\beta + v_i > 0\right) \\ &= 1 - P\left(v_i < -x'_i\beta\right) \\ &= 1 - \left[1 - P\left(v_i < x'_i\beta\right)\right] \\ &= P\left(v_i < x'_i\beta\right) \\ &= \Phi\left(x'_i\beta\right) \end{aligned} \tag{3.4}$$

Ordinal data models can be obtained by thresholding an underlying latent variable. Specifically, we consider that a continuous latent random variable z_i relies on a set of vector of explanatory variables x_i via the association $z_i = x_i'\beta + v_i$, but with the unique characteristic that the reported response $y_i \in (1, 2, ..., K)$ occur according to,

$$y_i = k \qquad if \qquad \tau_{k-1} < z_i \leq \tau_k \tag{3.5}$$

k denotes the ordinal variables. τ denotes cut points. $-\infty < \tau_0 < \tau_1 < ... < \tau_{K-1} < \tau_K < \infty$ are cut points that establish the categorization of the data into K ordered categories. The chance of reporting $y_i = k$, conditional on β and $\tau = (\tau_1, ..., \tau_{K-1})$ is given by

$$P(y_i = k|\beta,\ \tau) = \Phi\left(\tau_k - x'\beta_i\right) - \Phi\left(\tau_{k-1} - x'\beta_i\right) \tag{3.6}$$

Similarly, we demand restrictions on location and scale in order to find the parameters.

We restrict $\tau_0 = -\infty$, $\tau_K = \infty$ and assume $\tau_1 = 0$ to avoid the possibility of altering the distribution without changing the chance of observing y_i and dodge the handling of boundary parameters. The scale is still fixed to 1 for normalization.

According to Xin-She (2020), Classical estimation is based on maximum likelihood that demands maximization or minimization of the log-likelihood function and Bayesian paradigm that is achieved by Markov Chain Monte Carlo (MCMC) simulation methods such as Metropolis Hasting and Gibbs sampling.

Consider a set of responses $y = (y_1, ..., y_n)'$ that originate from some statistical model with sampling density $f(y|\theta)$ given by parameter vector θ. The density $f(y|\theta)$ presents a mathematical description of the probabilistic phenomenon that creates the response variable y given θ, it is called the data generating process. When $f(y|\theta)$ is looked as a function of the parameter vector θ given the sample y, it is called the likelihood function.

The value of θ that maximizes the log-likelihood function is referred to as maximum likelihood estimator.

$$\hat{\theta}_{MLE} = \arg \max_{\theta} \log \left(f(y|\theta)\right) \tag{3.7}$$

It is the value of θ that makes the observed sample y as "likely" as possible within the limits of the proposed data generating process. Particularly, it is acknowledged that under mild regularity conditions, the estimator $\hat{\theta}_{MLE}$ is consistent and asymptotically normally distributed.

Consistency indicates that as $n \to \infty$, the probability limit of $\hat{\theta}_{MLE}$ is the true value θ_0 , i.e. $.p\lim \hat{\theta}_{MLE} = \theta_0$. Asymptotic normality means that in large samples, as $n \to \infty$, where V is the Hessian matrix of the log-likelihood evaluated at θ_0 and $\hat{\theta}_{MLE} \sim N\left(\theta_0, V^{-1}\right)$ the expectation is taken with respect to $f(y|\theta_0)$. Since it is difficult to evaluate this expectation, it is normal to approximate V by the observed Hessian matrix $V = -\frac{\partial^2 Inf(y|\theta)}{\partial\theta'\partial\theta}$ which is evaluated at the maximum likelihood value $\theta = \hat{\theta}_{MLE}$. The standard errors of the estimated parameters are obtained by finding the square root of the diagonal entries of V^{-1}. The standard errors are then used in creating the confidence intervals and test of hypothesis. For the binary data models, the likelihood function can be written as,

$$\begin{aligned} f(y|\beta) &= P(y_1, y_2, ..., y_n|\beta) \\ &= \prod_{i=1}^{n} f(y_i|\beta) \\ &= \prod_{i:y_i=1}^{n} \Phi\left(x'_i\beta\right) \prod_{i:y_i=0}^{n} \left[1-\Phi\left(x'_i\beta\right)\right] \\ &= \prod_{i=1}^{n} \left[\Phi\left(x'_i\beta\right)\right]^{y_i=1} \prod_{i=1}^{n} \left[1-\Phi\left(x'_i\beta\right)\right]^{y_i=0} \end{aligned} \tag{3.8}$$

where the second equation is obtained by assuming observations are independent and the final equation is basically a convenient form for outlining the corresponding probability.

The log-likelihood function is given by

$$logf(y|\beta) = \sum_{i=1}^{n} y_i log(\Phi\left(x'_i\beta\right)) + (1-y_i)\, log\left(1-\Phi\left(x'_i\beta\right)\right) \tag{3.9}$$

The maximization is carried out iteratively using standard hill climbing algorithms such as Newton-Raphson because the first-order condition for maximization does not admit an explicit analytical solution even though the log-likelihood is typically well behaved (unimodal and concave) in this class of models. The score vector for the probit model is given by,

$$\frac{\partial logf(y|\beta)}{\partial\beta} = \sum_{i=1}^{n} y_i \frac{\phi(x'_i\beta)}{\Phi(x'_i\beta)} - (1-y_i)\frac{\phi(x'_i\beta)}{1-\Phi(x'_i\beta)} x_i \tag{3.10}$$

where ϕ is a probability density function and Φ is a cumulative distribution function. The second derivative of probit model is given by,

$$\begin{aligned} \frac{\partial^2 logf(y|\beta)}{\partial\beta\partial\beta'} = -\sum_{i=1}^{n} \phi\left(x'_i\beta\right) \Bigg[& y_i \frac{\phi(x'_i\beta) + x'_i\beta\Phi(x'_i\beta)}{\Phi(x'_i\beta)^2} \\ & + (1-y_i)\frac{\phi(x'_i\beta) - x'_i\beta\,(1-\Phi(x'_i\beta))}{(1-\Phi(x'_i\beta))^2} \Bigg] x_i x'_i \end{aligned} \tag{3.11}$$

The Newton-Raphson iteration is given by,

$$\tilde{\beta}_{n+1} = \tilde{\beta}_n - \left[\frac{\partial^2 l}{\partial\beta\partial\beta'}\right]^{-1}_{\beta=\tilde{\beta}} \left[\frac{\partial l}{\partial\beta}\right]_{\beta=\tilde{\beta}} \tag{3.12}$$

$\tilde{\beta}_n$ is the n^{th} round estimate and the Hessian and score vectors are evaluated at this estimate. The stopping criteria is specified by $\left|\tilde{\beta}_{n+1} - \tilde{\beta}_n\right| < \varepsilon$. From our Maximum Likelihood theorem, we know that,

$$\sqrt{N\left(\tilde{\beta}_{ML} - \beta\right)} \xrightarrow{asy} N\left(0, \left(-E\left(\frac{\partial^2 l}{\partial\beta\partial\beta'}\right)^{-1}\right)\right) \tag{3.13}$$

For finite samples, the asymptotic distribution of $\tilde{\beta}_{ML}$ can be approximated by
$N\left(\beta, -\left(-\frac{\partial^2 l}{\partial\beta\partial\beta'}\right)^{-1}_{\beta=\beta_{ML}}\right)$.
The likelihood function for ordinal outcomes under the assumption of independent sampling is given by,

$$f(y|\beta,\tau) = P(y_1, y_2, ..., y_n|\beta) = \prod_{i=1}^{n} f(y_i|\beta,\tau) \tag{3.14}$$

$$= \prod_{i=1}^{n} \left[\Phi\left(\tau_k - x'_i\beta\right) - \Phi\left(\tau_{k-1} - x'_i\beta\right)\right]^{y_i=k} \tag{3.15}$$

To minimise the computational complexities related to constrained optimization, it is helpful to reparametrize the problem in order to eliminate the constraints. This can be done by transforming the cut points τ so as to eliminate the ordering constraint by one-to-one mapping $\delta_k = log(\tau_k - \tau_{k-1})$ and $2 \leq \mathrm{k} \leq \mathrm{K} - 1$ and then amending the likelihood function to be a function of β and δ.

3.3 Zero Inflated Ordered Probit Model

The model was proposed by Harris and Zhao (2007). They incorporated excess zero by considering a latent variable model (similar to a binary selection model):

$$s_i^* = w_i'\gamma + \mu_i \tag{3.16}$$

where s_i^* denotes a latent variable, w_i denotes $p \times 1$ vector of covariates and γ denotes related vector of coefficients. The binary response is given by

$$s_i = I(s_i{}^* > 0) \tag{3.17}$$

s_i denotes the observed binary variable. Where $I(s_i{}^* > 0) = 1$ if $s_i{}^* > 0$, and 0 otherwise. In scheme one, $s_i = 1$ or $s_i{}^* > 0$ for participants (e.g., smokers) while in scheme zero, $s_i = 0$ or $s_i{}^* \leq 0$ for non-participants. In case of the error term, $s_i = 0$ implies that $\mu_i \leq -w_i{}'\gamma$. Let $\tilde{y}_i^*$ and $\tilde{y}_i$ denotes the latent variable and observed ordinal variable respectively. In setting of the zero-inflation model, the observed response y_i assumes the form

$$y_i = s_i \tilde{y}_i \tag{3.18}$$

We report $y_i = 0$ when either the respondent is a non-participant ($s_i = 0$) or the respondent is a zero consumer ($s_i = 1$ and $\tilde{y}_i = 0$). Likewise, we report positive outcome (consumption) when the respondent is a positive consumer ($s_i = 1$ and $\tilde{y}_i >$ 0). Assume the error terms are independently distributed. The zero-inflated ordered multinomial distribution, say $P(y_i)$, arises as a mixture of a degenerate distribution at zero and the assumed distribution of the response variable $\tilde{y}_i$ as follows:

$$f_1\left(\tilde{y}_i^*, \tilde{y}_i, s_i{}^*, s_i | x_i, w_i, \Psi\right) = \begin{cases} P\left(s_i = 0\right) + P\left(s_i = 1\right) P\left(\tilde{y}_i = 0\right) & \text{if } k = 0, \\ P\left(s_i = 1\right) P\left(\tilde{y}_i = k\right) & \text{if k = 1, 2, ..., K} \end{cases} \tag{3.19}$$

The parameter space is given by $\Psi = (\beta, \gamma, \alpha, \Omega_0)$. The likelihood function based on N independent responses is given by

$$\begin{aligned} L\left(\tilde{y}_i^*,\ \tilde{y}_i,\ s^*,\ s | x,\ w,\ \Psi\right) = & \prod_{i=1}^{N} \prod_{k=0}^{K} \left[P\left(s_i = 0\right) + P\left(s_i = 1, \tilde{y}_i = 0\right)\right]^{d_{ik}} \\ & \times \prod_{i=1}^{N} \prod_{k>0}^{K} \left[P\left(s_i = 1, \tilde{y}_i = k\right)\right]^{d_{ik}} \end{aligned} \tag{3.20}$$

d_{ik} is an indicator variable where $d_{ik} = 1$ if the respondent i picks outcome k and $d_{ik} = 0$ otherwise.

Separate alternatives of the description of the joint distribution of $(\varepsilon_i,\ \mu_i)$ produces different zero inflated ordered response models. If the error terms are correlated, we obtain ZIOPC and in case they are uncorrelated, we obtain independent ZIOP model.

3.4 Autoregressive Ordered Probit Model

AOP model was proposed by Muller and Czado (2005) and incorporated a lagged latent dependence in OP model. Let us consider y_t as an observed ordinal response at $t = 1,\ 2,\ ...,\ T$ where y_t assumes only K distinct status, and a $r+1$ dimensional vectors $x_t = (1,\ x_{t1},\ x_{t2},\ ...,\ x_{ir})'$ of independent variables for every period. Let ${y_t}^*$ be an unobserved autoregressive dependent variable obtained from the observed ordinal response y_t by thresholding. It is represented by,

$$y_t = k \Leftrightarrow y_t^* \in (\tau_{k-1},\ \tau_k]\ k = 1,\ 2,\ ...,\ K \tag{3.21}$$

$$y_t^* = x_t'\varphi + \ \phi y_{t-1}^*\ +\ \upsilon_t^*\ t = 1,\ 2,\ ...,\ T \tag{3.22}$$

where $-\infty < \tau_0 < \tau_1 < ... < \tau_{k-1} < \tau_k = \infty$ are unknown cut points, $\varphi = (\varphi_1,\ \varphi_2,\ ...,\ \varphi_p)^T$ is a transposed column matrix of unknown regression coefficients of the auto regression probit model and ϕ is the autocorrelation coefficient. We let $\upsilon_t^* \sim N\left(0, \delta^2\right)\ i.i.d.$

3.5 Dynamic Panel Ordered Probit with Random Effects

The dynamic panel ordered probit model proposed by Hasegawa (2009) could be used to model an ordinal dependent variable, e.g. $y_{it} = 1,\ 2,\ ...,$ K. Let $y_{it}^* \in\ (-\infty,\ \infty)$ be the underlying latent variable for an individual i at period t for $i = 1,\ 2,\ ...,\ N$. The model is given by,

$$y_{it} = k \tag{3.23}$$

if $y_{it}^{*} \in (\tau_{k-1},\ \tau_{k}]$ where τ_k is a cut point and k is an ordinal observation. The latent variable is considered to be established by the subsequent linear models

$$y_{it}{}^{*} = \phi y_{it-1}^{*} + \gamma' x_{it} + \kappa_i + \varepsilon_{it},\ t = 1,\ 2,\ 3,\ ...,\ T \tag{3.24}$$

$$y_{i0}{}^{*} = \gamma' x_{i1} + \kappa_i + \varepsilon_{i0},\ t = 0 \tag{3.25}$$

where $x_i = (x_{i1},\ x_{i2},\ ...,\ x_{ip})$ is the time-invariant and time-variant covariates, κ_i is an unobserved heterogeneities. ε_{it} is a time and individual-specific error term which is considered to be normally distributed and independent across respondents and times and indepedent of κ_i. κ_i is considered to be normally distributed with a mean of zero and constant variance, σ^2 , and uncorrelated with ε_{it} for all t. ε_{it} is assumed to be strictly exogenous, that is, ε_{it} are independent of ε_{is} for all t and s.

Conditioned on the unobserved heterogeneities κ_i, the response on y_{it}, are considered to be independent. Then, the contribution to the likelihood for respondent i, conditional on the explanatory variables and the unobserved heterogeneities, would be the joint probability.

$$\begin{aligned} p(y_{it} = k | x_{it}, \kappa_i) = & \prod_{i=1}^{N}\prod_{t=0}\prod_{k=0}^{K} \Phi\left(\tau_k - \gamma'_2 x_{i1} - \kappa_i\right) - \Phi\left(\tau_{k-1} - \gamma'_2 x_{i1} - \kappa_i\right) \\ & \prod_{i=1}^{N}\prod_{t=1}^{T}\prod_{k=0}^{K} \left[\Phi\left(\tau_k - \phi y_{it-1}^{*} - \gamma'_2 x_{it} - \kappa_i\right) - \right. \\ & \left. \Phi\left(\tau_{k-1} - \phi y_{it-1}^{*} - \gamma'_2 x_{it} - \kappa_i\right)\right] \end{aligned} \tag{3.26}$$

where $\Phi(.)$ is the normal distribution function.

3.6 Gauss-Hermite Quadrature Approximation (GHQ)

The ZIOP model has an intractable likelihood function. Approximations such as Gauss-Hermite quadrature need to be utilized to compute the multidimensioanl integrals. Maximum likelihood estimation of dynamic latent variable models requires evaluation of integrals that are intractable. Numerical approximations provides a fix to this difficulty.

The probabilists Hermite polynomials is given by,

$$H_K(u) = (-1)^K \exp\left\{\frac{u^2}{2}\right\} \frac{d^K}{du^K} \exp\left\{\frac{u^2}{2}\right\} \tag{3.27}$$

with weights $w_k = \frac{2^{K-1}K!\sqrt{\pi}}{K^2[H_{K-1}(u_k)]^2}$ and nodes μ_k for $k = 1, 2, ..., K$ are the roots of probabilists Hermite polynomials.

The μ_k is the k^{th} zero of the K^{th} order Hermite polynomial $H_K(\mu_k)$. It is important to note that the nodes are symmetrically distributed around zero. The approximation is exact if $f(u)$ is a polynomial of order $2K-1$ at the maximum. The tables for the values of μ_k and w_k are available in Abramovitz and Stegun (1964).

Gaussian quadrature is an influential tool for computing the intractable integrals and is based on weights and nodes. Gauss-Hermite quadrature (GHQ) is used for numerical approximation of integrals with Gaussian kernels. The integrand $g(z)$ can be factored as $f(z)\exp(-z^2)$. Here $\exp(-z^2)$ is called the weight function. When the Gaussian density is not the factor of the integrand, a linear transformation can be made to convert the Gaussian factor into the form $\exp(-z^2)$.

Consider the following integral

$$\int f(z)\exp(-z^2)\,dz \tag{3.28}$$

GHQ approximates this integral as a finite weighted sum

$$\int f(z)\exp(-z^2)\,dz \approx \sum_{k=1}^{K} w_k f(\mu_k) \tag{3.29}$$

Classical GHQ approximation can be used when the weight function in the integrand takes the form $\exp(-z^2)$. When the weight function is a Gaussian density with mean μ and variance σ^2 , that is, $\varphi(z;\mu,\sigma^2)$, a linear transformation to z will be used to convert the weight function into the form $\exp(-z^2)$. This technique shifts the location from zero to the mean of the weight function and changes the scale by incorporating the dispersion parameter into the values of the new nodes. After applying this transformation the nodes become symmetric about the mean μ of the weight function.

Let us assume that the weight function is a Gaussian density with μ and σ^2. We define the integrand as $g(z) = f(z)\,\varphi\left(z;\mu,\sigma^2\right)$. Consider the integral of the form

$$\int g(z)dz = \int f(z)\,\varphi\left(z;\mu,\sigma^2\right)dz \tag{3.30}$$

or

$$\int g(z)dz = \frac{1}{\sqrt{2\pi\sigma^2}}\int f(z)\exp\left\{-\frac{(z-\mu)^2}{2\sigma^2}\right\}dz \tag{3.31}$$

By variable transformation $u = \frac{z-\mu}{2\sigma^2}$, we get $z = \mu + \sqrt{2\sigma^2}u$.The above integral can be rewritten and approximated as,

$$\int \frac{1}{\sqrt{\pi}} f\left(\mu + \sqrt{2\sigma^2}u\right)\exp\left(-u^2\right)du \approx \sum_{k=1}^{K}\frac{1}{\sqrt{\pi}} w_k^* f\left(\mu_k^*\right) \tag{3.32}$$

Where $\mu_k^* = \mu + \sqrt{2\sigma^2}u_k$ are the shifted and scaled nodes and $w_k^* = {}^{w_k}/_{\sqrt{\pi}}$ the transformed weights. Locating the optimal number of nodes and weighs points can be achieved numerically. To achieve this, one can begin with a few number of nodes and weights points and raise them and note if it significantly affects the results, and repeat this process until there is insignificant change. However, increasing number of nodes and weights points also escalate the computing period. The accuracy of the Gauss–Hermite approximation relies on the chosen number of nodes and weights points. Ideally, number of nodes and weighs points is established by scrutinizing the convergence behaviour of $\sum_{k=1}^{K}\frac{1}{\sqrt{\pi}}w_k f(z_k)$ when the number of quadrature points is increased.

3.7 Bayesian inference

Sometime, GHQ algorithm results in high computational cost. The cost of direct computing the multidimensional integrals can be evaded by Bayesian approach. Bayesian approach doesn't need a large sample to ensure the adequacy of asymptotic

approximations. The approach possess the finite sample properties and are consistent and asymptotically efficient under mild conditions.
It relies on Bayes' theorem given by,

$$f(\vartheta|y) = \frac{f(y|\vartheta)\,\pi(\vartheta)}{\int f(y|\vartheta)\,\pi(\vartheta)\,d\vartheta} \propto f(y|\vartheta)\,\pi(\vartheta) \tag{3.33}$$

and inference depends on the posterior density $f(\vartheta|y)$ that is proportional to the product of the likelihood function and prior density . $\int f(y|\vartheta)\,\pi(\vartheta)\,d\vartheta$ is referred to as the normalizing constant and is free of ϑ, only a function of data.
Prior distribution indicates our thought about the ambiguity in ϑ before we collect any data while the posterior distribution indicates the ambiguity in ϑ after we collect the dataset y. The Prior distribution may be informative, diffused or completely non-informative. The non-informative is popular as it allow the data to dictate the analysis. Once we obtain the posterior distribution of ϑ, we can investigate its properties like measure of central tendencies such as posterior mean and measure of variability such as variance by analyzing the function $f(\vartheta|y)$. In many situations, notably when ϑ is a vector of parameters, it complex to isolate the properties of ϑ due to hurdles in solving the problem numerically. One of the possibility is to pursue MCMC algorithm outlined below.

3.8 Monte Carlo Methods

It represents a simulation-based approximation to compute intractable integrals. Its foundation originate from the law of large numbers that states as follows; If X is a random variable with $E(X) < \infty$ and $x_1,\ x_2,\ \ldots,\ x_h$ are i.i.d draws from the distribution of X , then as $h \to \infty$

$$\frac{1}{h}\sum_{i=1}^{h} x_i \to E(X) \text{ with probability } 1 \tag{3.34}$$

Hence, if we require to evaluate any integral of the form $\int f(y|\vartheta)\,\pi(\vartheta)\,d\vartheta$ for a recognizable integrable function, we can in turn simulate a huge number of (i.i.d.) observations $\vartheta_1,\ \vartheta_2,\ \ldots,\ \vartheta_h$ from posterior density of ϑ, compute the function π at those h points and approximate this integral with $\frac{1}{h}\sum_{i=1}^{h}\pi(\vartheta_i)$. The methods rely on the ability to create a huge number of observations from the target posterior of ϑ. One can also draw from the joint posterior by applying conditional and marginal draws in succession.

A sequence of a random variable $\left\{x^{(0)},\, x^{(1)},\, x^{(2)},\, ...\right\}$ is a Markov Chain if the conditional distribution of $x^{(n)}$, given $x^{(0)},\, x^{(1)},\, x^{(2)},\, ...,\, x^{(n-1)}$ merely relies on $x^{(n-1)}$. If $x^{(0)} \sim f_0\left(x^{(0)}\right)$, then

$$f_1\left(x^{(1)}\right) = \int q\left(x^{(1)}|x^{(0)}\right) f_0\left(x^{(0)}\right) dx^{(0)} \tag{3.35}$$

Here $q\left(x^{(1)}|x^{(0)}\right)$ is called transition kernel of the Markov chain. In general,

$$f_t\left(x^{(t)}\right) = \int q\left(x^{(t)}|x^{(t-1)}\right) f_{t-1}\left(x^{(t-1)}\right) dx^{(t-1)} \tag{3.36}$$

If $p_s(x)$ is a probability density function such that $x^t \sim p_s \Rightarrow x^{t+1} \sim p_s$ then $p_s(x)$ is the stationary distribution for the Chain. Obviously, the form of p_s (if it occur) relies on the form of q. If we simulate $x^{(0)} \sim p_s$, all subsequent steps will yield correlated samples from p.

We call a distribution $p_L(x)$ to be the limiting distribution of a Markov chain if,

$$p_L(x) = \lim_{t\to\infty} p\left(X^{(t)} \in A|X^{(0)} = x^{(0)}\right) \tag{3.37}$$

does not relies on the first state $x^{(0)}$. Limiting distribution does not always exists. We label the chain as irreducible, if there is a route to move from every state to every other state. We label the chain as aperiodic if for any two states a and b, the gcd of all path lengths that go from a to b is 1. We label the chain as positive recurrent, if beginning from any state, the average time to resume to that state is finite.

For a positive recurrent, irreducible and aperiodic chain, that is ergodic Markov chain, there exists a limiting distribution that is also its distinct stationary distribution. This indicates that if we can discover an ergodic Markov chain with stationary distribution $f(\vartheta|y)$, the marginal distribution of draws from that ergodic chain will converge to a limiting distribution distribution that is identical to the stationary distribution $f(\vartheta|y)$, regardless of the distribution of the first parameter vector $\vartheta^{(0)}$. The sampling scheme is called MCMC algorithm.

Suppose $q\left(\vartheta^{(t+1)}|\vartheta^{(t)}\right)$ be a transition kernel with stationary distribution $f(\vartheta|y)$. If we draw $\vartheta^{(0)}$ from preferred distribution and keep drawing $\vartheta^{(1)}, \vartheta^{(2)}$ as:

$$\vartheta^{(0)}\ q\left(\vartheta^{(1)}|\vartheta^{(0)}\right)\ \vartheta^{(1)}\ q\left(\vartheta^{(2)}|\vartheta^{(1)}\right)\ \vartheta^{(2)} \to ... \tag{3.38}$$

then, after a huge number of draws N are completed, $\vartheta^{(N+1)}$, $\vartheta^{(N+2)}$, ... can be approximated as correlated samples from $f(\vartheta|y)$. Thus, we need to dispense with number of initial draws, called burn-in period in an MCMC. The initial phase of the algorithm may be biased by the first values, and are therefore usually abandoned before further analysis.

Often, there is autocorrelation among the draws in each parameter sequence, while high correlation among draws in each parameter sequence makes the convergence of the sequence slow, some authors such as Plummer *et al.* (2005) and Zuur *et al.* (2002) recommended using thinning to reduce autocorrelation. This is done by only picking draws of the chain at a particular point d such as $\vartheta^{(N+1)}$, $\vartheta^{(N+d+1)}$, $\vartheta^{(N+2d+1)}$, ... A larger value of d produces a weaker correlation between the sequential observations. An MCMC algorithm normally utilizes the burn-in and thinning so that the remaining samples estimate as much as possible a set of independent draws from $f(\vartheta|y)$.

3.9 Gibbs sampling

Gibbs sampling is an MCMC method for simulation from a distribution when its full conditional densities have recognizable form. Gelfand and Smith (1990) reviewed Gibbs sampling and revealed its great potential in computing Bayesian posterior densities for a diverse structured models. Assume that there are k parameter blocks ϑ_1, ϑ_2, ..., ϑ_k with joint density $\pi(\vartheta_1, \vartheta_2, ..., \vartheta_k|y)$. We set the initial values for the parameter as ϑ_1^o, ϑ_2^o, ..., ϑ_k^o. Good starting values can be maximum likelihood estimates. At the h^{th} iteration, each element of ϑ_i^h, $i = 1, 2, ..., k$ is revised by sampling from the complete conditional distributions:

$$\begin{aligned}
\vartheta_1^{(h)} &\sim \pi\left(\vartheta_1|\vartheta_2^{(h-1)}, \vartheta_3^{(h-1)}, ..., \vartheta_k^{(h-1)}, y\right) \\
\vartheta_2^{(h)} &\sim \pi\left(\vartheta_2|\vartheta_1^{(h-1)}, \vartheta_3^{(h-1)}, ..., \vartheta_k^{(h-1)}, y\right) \\
\vartheta_k^{(h)} &\sim \pi\left(\vartheta_k|\vartheta_1^{(h-1)}, \vartheta_2^{(h-1)}, ..., \vartheta_{k-1}^{(h-1)}, y\right))
\end{aligned} \tag{3.39}$$

If we administer a huge number of iterations, the joint distribution of $\vartheta_i^{(h)} = \left(\vartheta_1^{(h)}, \vartheta_2^{(h)}, ..., \vartheta_k^{(h)}\right)$ approaches the posterior distribution. Under mild conditions, it can be shown that the Markov chain formed by the Gibbs sampler has a limiting invariant distribution that is the distribution of interest $\pi(\vartheta_1, \vartheta_2, ..., \vartheta_k|y)$. This implies that draws acquired by Gibbs sampling after the initial burn-in period, can be assumed as originating from $\pi(\vartheta_1, \vartheta_2, ..., \vartheta_k|y)$. We can run the algorithm as long as needed to reach the level of precision we want. The full conditionals are easy to sample from if the model is conjugate. However, the algorithm has some limitations. Sometime, it is impossible or impractical to obtain the conditional distributions for each of the random variables in the model or it may be that the conditional distributions are of unrecognizable form, and therefore there is no clear means to obtain samples from them. There are also situation in which Gibbs sampling will be very inefficient. For instance, the mixing of the Gibbs sampling chain might be very slow, implying that the algorithm may consume a long period of time exploring a local region with high density, and thus take very long to explore all regions with significant probability mass.

3.10 Metropolis-Hastings Algorithm

The algorithm was developed by Metropolis and Ulam (1949) and Hastings (1970) can be used when the Gibbs sampler fails, either because of poor convergence or the inability to sample from a known full conditional distributions. We locate a transition kernel q with stationary distribution same as $f(\vartheta|y)$ using the algorithm. Consider a case where a closed form expression is available for $f(\vartheta|y)$. We are unable to sample from it but, given a point, we can compute it up to a normalizing constant. The Metropolis-Hastings (MH) algorithm perform well in this case by proposing successive values of ϑ from a proposal distribution g that is completely recognizable and easy to draw from.

Given $\vartheta^{(t)}$, we can draw $\vartheta^{(propose)}$ from $g\left(\vartheta^{(propose)}|\vartheta^{(t)}\right)$. So, the recent state of ϑ acts as a parameter in g. Then, we compute an acceptance probability α_A provided by

$$\alpha_A\left(\vartheta^{(propose)}|\vartheta^{(t)}\right) = \min\left(1, \frac{f\left(\vartheta^{(propose)}|y\right)}{f\left(\vartheta^{(t)}|y\right)} \frac{g\left(\vartheta^{(t)}|\vartheta^{(propose)}\right)}{g\left(\vartheta^{(propose)}|\vartheta^{(t)}\right)}\right) \tag{3.40}$$

Finally, we set the next state of ϑ as:

$$\vartheta^{(t+1)} = \begin{cases} \vartheta^{(propose)} & \text{with probability } \alpha_A, \\ \vartheta^{(t)} & \text{with probability 1 - } \alpha_A, \end{cases} \tag{3.41}$$

The main feature of the algorithm is to guarantee a favorable rate of acceptance for the proposals. A useful proposal distribution will provide a value of α_A close to 1(so we accept what we propose most of the time). If a proposal distribution provides small values of α_A close to 0 most of the time, the Markov chain of α_A often gets stuck at present states and covers only a few states in a long time. In usages, it may be challenging to pick a proposal distribution with favorable rate of acceptance majority of the times.

There are two types of proposal distributions; namely symmetric and asymmetric proposal distributions in the literature. Symmetric proposal distribution exist when $q\left(\vartheta^{(t)}|\vartheta^{(t-1)}\right) = q\left(\vartheta^{(t-1)}|\vartheta^{(t)}\right)$. Straightforward choices of symmetric proposals include Gaussian distributions or Uniform distributions. For instance, if Gaussian proposal is available, then $\vartheta^{(proposed)} = \vartheta^{(t)} + N(0, \sigma)$. Because the probability density function for $N\left(\vartheta^{(proposed)} - \vartheta^{(t)}, 0, \sigma\right) = N\left(\vartheta^{(t)} - \vartheta^{(proposed)}, 0, \sigma\right)$, this is a symmetric proposal. A proposal distribution is a asymmetric distribution if $q\left(\vartheta^{(t)}|\vartheta^{(t-1)}\right) \neq q\left(\vartheta^{(t-1)}|\vartheta^{(t)}\right)$. This includes Student t distribution, Inverse Wishart distribution, log-normal distribution, distribution among others.

3.11 Type of proposals used in Metropolis-Hastings Algorithm

Random walk proposal: It uses symmetric distribution as a proposal density. Recommend a new state of ϑ from a distribution centered at its present state and a proposal variance. The proposed state is not automatically admitted in the posterior samples. They are admitted probabilistically based on the established probability α_A. This proposal distribution randomly upsets the present state of the chain, and then either admits or fail to admit the perturbed value. The acceptance function in this situation is given by

$$\alpha_A\left(\vartheta^{(propose)}|\vartheta^{(t)}\right) = \min\left(1, \frac{f\left(\vartheta^{(propose)}|y\right)}{f\left(\vartheta^{(t)}|y\right)}\right) \tag{3.42}$$

Since $g\left(\vartheta^{(propose)}|\vartheta^{(t)}\right) = g\left(\vartheta^{(t)}|\vartheta^{(propose)}\right)$.

Independent proposal: It is recommended when the proposal distribution is asymmetrical such as Chi-square distribution and this permit us to accommodate a specific constraints in our models. These constraints include limiting the correlation between -1 and 1, non-negative variance etc. In this case, the present state of ϑ is not used to propose a new state. Hence, we denote $g\left(\vartheta^{(propose)}|\vartheta^{(t)}\right) = g\left(\vartheta^{(propose)}\right)$, free of $\vartheta^{(t)}$. The acceptance function in this situation is given by :

$$\alpha_A\left(\vartheta^{(propose)}|\vartheta^{(t)}\right) = \min\left(1, \frac{f\left(\vartheta^{(propose)}|y\right)}{f\left(\vartheta^{(t)}|y\right)} \frac{g\left(\vartheta^{(t)}\right)}{g\left(\vartheta^{(propose)}\right)}\right) \tag{3.43}$$

The proposal is not connected to the present values of ϑ. Therefore, we can recommend more scattered values across the domain.

3.12 Single Component Metropolis-Hastings

The Metropolis Hasting algorithm uses a single step to updates the whole parameter vector. Therefore, a q-dimensional parameter vector demand a q-dimensional proposal distributions. Nevertheless, at times it is easier to update each parameter individually, in an algorithm referred to as single component Metropolis-Hastings (SCMH).

Let $\vartheta^{(n)} = \left(\vartheta_1^{(n)}, \vartheta_2^{(n)}, \ldots, \vartheta_p^{(n)}\right)$ be a p-dimensional vector denoting our relevant parameters at iteration n. Define $\vartheta_{-1} = (\vartheta_1, \ldots, \vartheta_{i-1}, \vartheta_{i+1}, \ldots, \vartheta_p)$ as the vector ϑ with the i^{th} parameter eliminated.

The algorithm updates each parameter in ϑ individually. What is required to update the i^{th} parameter of ϑ^n is a proposal distribution $q\left(y|\vartheta_i^{(n)}, \vartheta_{i-1}^{(n)}\right)$ for proposing a fresh value of ϑ_i given the present value of ϑ_i and all the other parameters.

Let $X = (X_1, X_2, \ldots, \mathrm{X}_p)$. There are p coordinate wise updates in the algorithm represent each stage.

Let $X_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, \mathrm{X}_p)$ and $X_{-i}^n = \left(X_1^{n+1}, \ldots, X_{i-1}^{n+1}, X_{i+1}^{n+1}, \ldots, \mathrm{X}_p^n\right)$. $i = 1, 2, \ldots, p$. X_{-i}^n is the update at stage n, where first $i-1$ coordinates are updated to their values at step $n+1$ and the coordinates at positions $i+1, i+2, \ldots, p$ remain at step n . Define $q_i\left(Y_i|X_i^n, X_{-i}^n\right)$ as the proposal distribution that creates proposals for i^{th} coordinate only. Then, the algorithm progresses as follows:

Create a candidate $Y_i \sim q_i\left(Y_i|X_i^n, X_{-i}^n\right)$.

Admit the candidate Y_i with probability

$$\alpha\left(X_i^n, X_{-i}^n, Y_i\right) = \min\left(1, \frac{\pi\left(Y_i|X_{-i}^n\right) q_i\left(X_i^n|Y_i, X_{-i}^n\right)}{\pi\left(X_i|X_{-i}^n\right) q_i\left(Y_i|X_i, X_{-i}^n\right)}\right) \tag{3.44}$$

as $X_i^{n+1} = Y_i$. Otherwise $X_i^{n+1} = X_i^n$. Proceed to the next coordinate.

Accept/Reject a proposal: Ultimately, we admit a newly proposed value with the acceptance probability α, otherwise it not admitted. The min operator in the function ensure the probability α does not exceed 1. Practically, we sample a random number uniformly between 0 and 1, and if this random number is less than α, we admit the newly proposed value; otherwise it is discarded.

However, we need to monitor the acceptance rates, that is, the times the algorithm rejected the newly proposed values. When the variance of the proposed distribution is small, the newly proposed values will be very close to the present state resulting into a high α that raises the acceptance rate. This will produce small moves within the domain and significantly raise the time it take to traverse the entire domain ϑ. This will leads to suboptimal exploration of the parameter space. When the variance of the proposed distribution is large, the newly proposed values will be very far from the current state resulting into a low α that lowers the acceptance rate. This will produce large moves within the domain and significantly raise the time it take to traverse the entire domain ϑ. This will leads to suboptimal exploration of the parameter space. Consequently, we require to balance the moves within the domain in order to efficiently traverse the whole domain of ϑ. The acceptance rates within 20% - 30% are optimal for typical applications. We can either lower or raise the acceptance rates by increase or decrease variances respectively. In Gibbs sampling, acceptance rates are always 100% because all the newly proposed value are admitted. It is a unique case of Metropolis-Hastings algorithm where proposal distributions are replaced by posterior conditional distributions.

3.13 Metropolis Hasting within Gibb sampling

Hierarchical models are normally designed as products of conditional distributions and therefore the Gibbs sampler is prevalent in Bayesian modelling. When it become infeasible to apply Gibb sampling, we opt for Metropolis-Hastings algorithm instead- this is referred to as Metropolis within Gibbs sampling.

The Probit model produces the following likelihood function,

$$L(\beta|y_i) = \prod_{i=1}^{n} \left[\Phi\left(x'_i\beta\right)\right]^{y_i=1} \prod_{i=1}^{n} \left[1-\Phi\left(x'_i\beta\right)\right]^{y_i=0} \tag{3.45}$$

A multivariate normal distribution is normally chosen as a prior for β with mean vector b and covariance matrix B, that is, $\beta \sim N(b,B)$. This posterior distribution is given by

$$L(\beta|y_i) = \pi(\beta) \prod_{i=1}^{n} \left[\Phi\left(x'_i\beta\right)\right]^{y_i=1} \prod_{i=1}^{n} \left[1-\Phi\left(x'_i\beta\right)\right]^{y_i=0} \tag{3.46}$$

This is an incomplete conditional distribution and will be computationally complicated due to unclosed form. The complete conditional distributions are the distribution of one parameter in the model given all other parameters in the model and the data, y.

Similarly, the ordinal data model will also produce an incomplete conditional distribution that cannot be sampled directly. If the prior distribution is picked from a recognizable distributions such as normal distribution, the data enroll into the likelihood $f(y|\vartheta)$ in such a way that the posterior $\pi(\vartheta|y)$ have unknown distribution. The Metropolis-Hastings algorithm is well adapted for sampling from such non-standard distributions.

CHAPTER FOUR

ZERO INFLATION DYNAMIC PANEL ORDERED PROBIT MODEL BASED ON MAXIMUM LIKELIHOOD

4.1 Introduction

This chapter introduces a ZIDPOPI and ZIDPOPC models based on Rabe-Hesketh and Skrondal (2013) approach for initial condition, parametric distribution for unobserved individual effects, that is, bivariate normal distribution and assuming unknown cut points.

4.2 Zero Inflation Dynamic Panel Ordered Probit Model

Let y_{it}^{b} represent binary specification for either non-participation or participation for respondent i at period t where $i = 1, 2, ..., n$ and $t = 0, 1, 2, ..., T$. The respondents are sampled independently from the main population. The time $t = 0$ represents the first period.

The y_{it}^{b} represent a binary variable showing the partition between non-participation (no change, no symptoms), that is, $y_{it}^{b} = 0$ and participation (or change, symptoms), that is, $y_{it}^{b} = 1$ at time t. y_{it}^{b} is associated with latent variable y_{it}^{b*} through 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. A univariate, continuous, latent and autoregressive process y_{it}^{b*} is generated by the state dependence y_{it-1}^{b}, time invariant covariates w_i, time variant covariates x_{it} and error term u_{it}^{b} in a linear relationship.

The latent variables y_{it}^{b*} and y_{i0}^{b*} denotes the inclination for participation (change, symptoms) and are given in vector form by,

$$y_{it}^{b*} = \phi_1 y_{it-1}^{b} + \gamma_1' x_{it} + \beta_1' w_i + u_{it}^{b} \tag{4.1}$$

$$y_{i0}^{b*} = \gamma_{01}' x_{i0} + \beta_{01}' w_i + u_{i0}^{b} \tag{4.2}$$

ϕ_1, γ_{01} , γ_1 , β_{01} and β_1 are vectors of unknown coefficients. Where the response variable y_{it}^{b*} is latent and y_{i0}^{b*} denotes the first latent value of the process. $u_{it}^{b} = \kappa_{1i} + e_{it}^{b}$ is the composite error term. x_{it} and w_i are vectors of time variant covariates and time invariant covariates respectively assumed to be strictly exogenous covariates. Strictly exogenous implies that the covariates are independent from all previous, present and the upcoming values of e_{it}^{b}. e_{it}^{b} is considered to be strictly exogenous, that is, the x_{it} are indepedent from e_{is}^{b} for all t and s. The error terms of the model are considered normally distributed, that is, $e_{it}^{b} \sim iidN\left(0, \sigma_{eb}^{2}\right)$ 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 considered to be orthogonal (uncorrelated) to exogenous variables following the standard random-effects assumption. The association between two sequential error terms is given by $corr\left(u_{it}^{b},\, u_{is}^{b}\right) = \frac{\sigma_{\kappa_{1i}}^{2}}{\sigma_{\kappa_{1i}}^{2}+1}$ $(t,\ k = 1,\ 2, ...,\ T;\ t \neq s)$. $\sigma_{k_{1i}}^{2}$ denotes the variance of unobserved fixed effect in participation decision. To depict the state dependence, y_{it-1}^{b} is a vector of indicators for the respondent's status in the preceding wave and the model can be interpreted as a first order Markov process. If $\phi_1 \neq 0$, then the response y_{it-1}^{b} influences response in the next period t. The asymptotic properties was with fixed time T and the cross-sectional sample size, N, tending to infinity.

The probability of participation $P(y_{it}^{b} = 1)$ and $P(y_{i0}^{b} = 1)$ were given by,

$$P(y_{it}^{b} = 1|\phi_1 y_{it-1}^{b},\ \gamma_1' x_{it}',\ \beta_1' w_i,\ \kappa_{1i}) = P\left(y_{it}^{b*} > 0\right) = \Phi\left(\phi_1 y_{it-1}^{b} + \gamma_1{}' x_{it} + \beta'{}_1 w_i + \kappa_{1i}\right) \tag{4.3}$$

$$P(y_{i0}^{b} = 1|\ \gamma_{01}' x_{i0}',\ \beta_{01}' w_i,\ \kappa_{1i}) = P\left(y_{i0}^{b*} > 0\right) = \Phi\left(\gamma_{01}{}' x_{i0} + \beta'{}_{01} w_i + \kappa_{1i}\right) \tag{4.4}$$

and, by symmetry, for non-participation

$$P(y_{it}^{b} = 0|\phi_1 y_{it-1}^{b},\ \gamma_1' x_{it},\ \kappa_{1i}) = 1 - \Phi\left(\phi_1 y_{it-1}^{b} + \gamma'{}_1 x_{it} + \beta'{}_1 w_i + \kappa_{1i}\right) \tag{4.5}$$

$$P(y_{i0}^{b} = 0|\ \gamma_{01}' x_{i0},\ \beta_{01}' w_i,\ \kappa_{1i}) = 1 - \Phi\left(\gamma'{}_{01} x_{i0} + \beta'{}_{01} w_i + \kappa_{1i}\right) \tag{4.6}$$

where $\Phi\left(.\right)$ represents the cumulative normal distribution function.

Let y_{it}^{o} denotes ordinal discrete response for respondent i at period t where $i = 1,\ 2,\ ...,\ n$ and $t = 0,\ 1,\ 2,\ ...,\ T$. The respondents are sampled independently from the main population. The time $t = 0$ represents the first period. Conditional on $y_{it}^{b} = 1$, the consumption (change, severity of symptoms, nature of peace) levels are denoted by a categorical variable y_{it}^{o} ($y_{it}^{o} = 0, 1, 2, ..., K$) at time t that is created by an OP model through a second latent variable by means of thresholding. A univariate, continuous, latent and autoregressive process y_{it}^{o*} is generated by its state dependence y_{it-1}^{o}, z_{it} and u_{it}^{o} in a linear relationship. It can also be expressed in vector form as,

$$y_{it}^{o*} = \phi_2 y_{it-1}^{o} + \gamma_2' z_{it} + \beta_2' v_i + u_{it}^{o} \tag{4.7}$$

$$y_{i0}^{o*} = \gamma_{02}' z_{i0} + \beta_{02}' v_i + u_{i0}^{o} \tag{4.8}$$

ϕ_2 , γ_{02} , γ_2 , β_{02} and β_2 are vector of unknown coefficients. Where the response variable y_{it}^{o*} is latent and y_{i0}^{o*} denotes the first latent value of the process. $u_{it}^{o} = \kappa_{2i} + e_{it}^{o}$ 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. Strictly exogenous implies that the covariates are independent from all previous, present and the upcoming values of e_{it}^{o}. e_{it}^{o} is considered to be strictly exogenous, that is, the x_{it} are uncorrelated with e_{is}^{b} for all t and s . The error terms of the model are assumed as $e_{it}^{o} \sim iidN\left(0, \sigma_{eo}^{2}\right)$ and $\sigma_{eo}^{2} = 1$ for identification purpose. κ_{2i} 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 considered to be orthogonal to exogenous variables following the standard random-effects assumption. The inter class association between two sequential error terms is given by $corr\left(u_{it}^{o},\ u_{is}^{o}\ \right) = \frac{\sigma_{\kappa_{2i}}^{2}}{\sigma_{\kappa_{2i}}^{2}+1}$ $\left(t,\ k = 1,\ 2, ...,\ T;\ t \neq s\right)$. $\sigma_{k_{2i}}^{2}$ denotes the variance of unobserved fixed effect in participation decision. To depict the state dependence, y_{it-1}^{o} is a vector of indicators for the respondent's status in the preceding wave and the model can be interpreted as a first order Markov process. If $\phi_2 \neq 0$, then the outcome y_{it-1}^{o} influences outcome in the next period t. The asymptotic properties was with fixed T and the cross-sectional sample size, N, tending to infinity. The mapping between y_{it}^{o} and y_{it}^{o*} is given by,

$$y_{it}^{o}=\begin{cases}0 & \text{if } y_{it}^{b*}\leq 0,\\ 0 & \text{if } y_{it}^{b*}>0 \textit{ and } y_{it}^{o*}\leq 0,\\ k & \text{if } y_{it}^{b*}>0 \textit{ and } \tau_{k-1}<y_{it}^{o*}\leq \tau_{k} \quad k=1,\ 2,\ ...,\ \mathrm{K}-1\\ K & \text{if } y_{it}^{b*}>0 \textit{ and } y_{it}^{o*}>\tau_{K-1}\end{cases} \tag{4.9}$$

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

$$y_{it}^{o}=k \Leftrightarrow \tau_{k-1}<y_{it}^{o*}\leq \tau_{k} \quad k=0,\ 1,\ 2,\ ...,\ \mathrm{K} \tag{4.10}$$

where τ_k are cut points of ordinal response. To guarantee that the cumulative distribution function for y_{it}^{0} is well defined, we needed that $\tau_{k-1}<\ \tau_k\ \forall$k . We specify that $-\infty<\tau_{-1}<\tau_0<\tau_1<...<\tau_{K-1}<\tau_K<\infty$ where, $\tau_{-1}=-\infty$, $\tau_K=\infty$ and $\tau_0=0$ for identification purpose and to steer clear from handling of boundary parameters. Considering that e_{it}^{o} is Gaussian, the OP probabilities are given by

$$P\left(y_{it}^{o}\right)=\begin{cases}P\left(y_{it}^{o}=0|\phi_2 y_{it-1}^{o},\ \gamma'_2 z_{it},\ \beta'_2 v_i,\ \kappa_{2i},\ y_{it}^{b}=1\right)=\\ \Phi\left(-\phi_2 y_{it-1}^{o}-\beta'_2 v_i-\gamma'_2 z_{it}-\kappa_{2i}\right)\\ \text{if } k=0,\\ P\left(y_{it}^{o}=k|\phi_2 y_{it-1}^{o},\ \gamma'_2 z_{it},\ \beta'_2 v_i,\ \kappa_{2i},\ y_{it}^{b}=1\right)=\\ \Phi\left(\tau_k-\phi_2 y_{it-1}^{o}-\beta'_2 v_i-\gamma'_2 z_{it}-\kappa_{2i}\right)\\ -\Phi\left(\tau_{k-1}-\phi_2 y_{it-1}^{o}-\gamma'_2 z_{it}-\beta'_2 v_i-\kappa_{2i}\right)\\ \text{if k}=1,\ 2,\ ...,\ \mathrm{K}-1,\\ P\left(y_{it}^{o}=K|\phi_2 y_{it-1}^{o},\ \gamma'_2 z_{it},\ \beta'_2 v_i,\ \kappa_{2i},\ y_{it}^{b}=1\right)=\\ \Phi\left(\phi_2 y_{it-1}^{o}+\gamma'_2 z_{it}+\beta'_2 v_i+\kappa_{2i}-\tau_{K-1}\right)\\ \text{if } k=K\end{cases} \tag{4.11}$$

While y_{it}^{b} and y_{it}^{o} are not independently reported in terms of the zeros, they are reported through the criterion $y_{it} = y_{it}^{b} y_{it}^{o}$ as proposed by Harris and Zhao (2007). To report $y_{it} = 0$ outcome we need either $y_{it}^{b} = 0$ (the respondent is a nonparticipant) or jointly that $y_{it}^{o} = 0$ and $y_{it}^{b} = 1$ (the respondent is a participant but with zero consumption). To report a positive $y_{it}^{b} = 1$, we require jointly that the respondent is a participant $y_{it}^{b} = 1$ and that $y_{it}^{o} > 0$ (the respondent is a participant and with non-zero consumption).

Considering that e_{it}^{b} and e_{it}^{o} are Gaussian distributions, the full probabilities for y_{it} is given by

$$P\left(y_{it}\right) = \begin{cases} P(y_{it} = 0|x,\ z) = P(y_{it}^{b} = 0|x) + P(y_{it}^{b} = 1|x)P(y_{it}^{o} = 0|x,\ z,\ y_{it}^{b} = 1) \\ \text{if } k = 0, \\ P(y_{it} = k|x,\ z) = P(y_{it}^{b} = 1|x)P(y_{it}^{o} = k|x,\ z,\ y_{it}^{b} = 1) \\ \text{if k} = 1,\ 2,\ ...,\ \text{K} - 1, \\ P(y_{it} = K|x,\ z) = P(y_{it}^{b} = 1|x)P(y_{it}^{o} = K|x,\ z,\ y_{it}^{b} = 1) \\ \text{if } k = K \end{cases} \tag{4.12}$$

Or

$$P\left(y_{it}\right)=\begin{cases} P(y_{it}=0|x,\ z)=\left(1-\Phi\left(\phi_1 y^b_{it-1}+\gamma'_1 x_{it}+\beta'_2 w_i+\kappa_{1i}\right)\right) \\ +\Phi\left(\phi_1 y^b_{it-1}+\gamma'_1 x_{it}+\beta'_2 w_i+\kappa_{1i}\right)\Phi\left(-\phi_2 y^o_{it-1}-\gamma'_2 z'_{it}-\beta'_2 v_i-\kappa_{2i}\right) \\ \text{if } k=0, \\ P(y_{it}=k|x,\ z)=\Phi\left(\phi_1 y^b_{it-1}+\gamma'_1 x_{it}+\beta'_2 w_i+\kappa_{1i}\right)* \\ (\Phi\left(\tau_k-\phi_2 y^o_{it-1}-\gamma'_2 z'_{it}-\beta'_2 v_i-\kappa_{2i}\right)- \\ \Phi\left(\tau_{k-1}-\phi_2 y^o_{it-1}-\gamma'_2 z'_{it}-\beta'_2 v_i-\kappa_{2i}\right))\text{ if k}=1,\ 2,\ ...,\text{K}-1 \\ P(y_{it}=K|x,\ z)=\Phi\left(\phi_1 y^b_{it-1}+\gamma'_1 x_{it}+\beta'_2 w_i+\kappa_{1i}\right) \\ \Phi\left(\phi_2 y^o_{it-1}+\gamma'_2 z'_{it}+\beta'_2 v_i+\kappa_{2i}-\tau_{K-1}\right) \\ \text{if } k=K \end{cases} \tag{4.13}$$

In this way, the chances of a zero response has been "inflated" as it is a combination of the chances of "zero consumption" from the OP model plus the chance of "non-participation" from the binary probit model.

We can assume that e^b_{it} and e^o_{it} are correlated with a correlation coefficient equal to ρ_{ebeo} since they correspond to the same individual. Thus the probability of observing y_{it} becomes,

$$y_{it}=\begin{cases} 0 & \text{if } \left(y^{b*}_{it}\leq 0\right)\ or\ \left(y^{b*}_{it}>0\ and\ y^{o*}_{it}\leq 0\right), \\ k & \text{if } \left(y^{b*}_{it}>0\ and\ \tau_{k-1}<y^{o*}_{it}\leq\tau_k\right),\text{k}=1,\ 2,\ ...,\text{K}-1 \\ K & \text{if } \left(y^{b*}_{it}>0\ and\ y^{o*}_{it}>\tau_{K-1}\right). \end{cases} \tag{4.14}$$

that transform into the succeeding expressions for the probabilities:

$$P(y_{it}) = \begin{cases} P(y_{it}=0|x,\ z) = \left(1-\Phi\left(\phi_1 y^b_{it-1}+\gamma'_1 x_{it}+\beta'_2 w_i+\kappa_{1i}\right)\right) \\ +\Phi\left(\phi_1 y^b_{it-1}+\gamma'_1 x_{it}+\beta'_1 w_i+\kappa_{1i},\right. \\ \left.-\phi_2 y^o_{it-1}-\gamma'_2 z'_{it}-\beta'_2 v_i-\kappa_{2i},-\rho_{ebeo}\right) \\ \text{if } k=0, \\ P(y_{it}=k|x,\ z) = \\ \Phi\left(\phi_1 y^b_{it-1}+\gamma'_1 x_{it}+\beta'_1 w_i+\kappa_{1i},\right. \\ \left.-\tau_k-\phi_2 y^o_{it-1}-\gamma'_2 z'_{it}-\beta'_2 v_i-\kappa_{2i},-\rho_{ebeo}\right) \\ \Phi\left(\phi_1 y^b_{it-1}+\gamma'_1 x_{it}+\beta'_1 w_i+\kappa_{1i},\right. \\ \left.\tau_{k-1}-\phi_2 y^o_{it-1}-\gamma'_2 z'_{it}-\beta'_2 v_i-\kappa_{2i},-\rho_{ebeo}\right) \\ \text{if k} = 1,\ 2,\ ...,\ \text{K}-1, \\ P(y_{it}=K|x,\ z) = \\ \Phi\left(\phi_1 y^b_{it-1}+\gamma'_1 x_{it}+\beta'_1 w_i+\kappa_{1i},\right. \\ \left.\phi_2 y^o_{it-1}+\gamma'_2 z'_{it}+\beta'_2 v_i+\kappa_{2i}-\tau_{K-1},\ \rho_{ebeo}\right) \text{ if } k=K \end{cases} \tag{4.15}$$

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

The presence of unobserved heterogeneities and the treatment of the first observations are the challenges normally faced when dealing with dynamic panel data models. There are two framework in dynamic panel data models namely random or fixed effects framework. In the fixed effects framework κ_i is permitted to be associated with the covariates. In the random effects framework κ_i is independent of the covariates.

In our dynamic models, we assume the random effects framework, where κ_i is uncorrelated with the structural variables $y_{i,t-1}$ and x_{it}. The assumptions includes $\text{cov}(y_{i,t-1},\ \kappa_i) = 0$ and $\text{cov}(x_{it},\ \kappa_i) = 0$. A challenge faced in the framework is the popular initial conditions problem. We are unable to compute $P\left(y_{i1}^b,\ y_{i1}^o | y_{i0}^{b*},\ y_{io}^{o*},\ x_{i0},\ z_{i0}\right)$ due to the absence of the preceding state on y_{i1}^b and y_{i1}^o (that is y_{i0}^b and y_{i0}^o). Disregarding the initial conditions problem in the likelihood function also leads to overlooking the data generation process for the initial observation of the panel and treating them as exogenous or to be in equilibrium. This only happen if the individual random effects are degenerative. Otherwise, the initial conditions are explained by the individual random effects and disregarding them results into inconsistent estimates.

Heckman (1981a) and Wooldridge (2005) proposed approaches for dealing with the initial conditions problem. Heckman (1981a) suggested a model for the initial condition given the unobserved heterogeneities and the strictly exogenous covariates.This model considered to be identical to the model underlying the remaining process. Wooldridge (2005) suggested a model for unobserved heterogeneities conditional on the initial conditions and the strictly exogenous covariates. The two approaches produces consistent estimates under the assumption of correct description of the distribution of the errors.

Rabe-Hesketh and Skrondal (2013) 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:

$$\kappa_{1i} = h_1^b y_{i0}^{b*} + h_0^b + h_2^b \bar{x}_i + h_3^b x_{i0} + \delta_{1i} \tag{4.16}$$

$$\kappa_{2i} = h_1^o y_{i0}^{0*} + h_0^o + h_2^o \bar{z}_i + h_3^o z_{i0} + \delta_{2i} \tag{4.17}$$

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

$$\kappa_{1i}|\left(y_{io}^{b*},\, \bar{x}_i\right) \sim N\left(h_1^b y_{i0}^{b*} + h_0^b + h_2^b \bar{x}_i + h_3^b x_{i0}, \sigma_1^2\right) \tag{4.18}$$

$$\kappa_{2i}|\left(y_{i0}^{o*},\, \bar{z}_i\right) \sim N\left(h_1^o y_{i0}^{o*} + h_0^o + h_2^o \bar{z}_i + h_3^o z_{i0}, \sigma_2^2\right) \tag{4.19}$$

where $\delta_{1i}|\left(y_{i0}^{b*},\, \bar{x}_i\right) \sim N\left(0,\, \sigma_1^2\right)$ and $\delta_{2i}|\left(y_{i0}^{o*},\, \bar{z}_i\right) \sim N\left(0,\, \sigma_2^2\right)$.

$h_0^b,\, h_1^b,\, h_2^b,\, h_3^b,\, h_0^o,\, h_1^o$, h_2^o and h_3^o represents parameters to be estimated. δ_{1i} and δ_{2i} are uncorrelated with $\left(y_{io}^b,\, \bar{x}_i\right)$ and $\left(y_{i0}^o,\, \bar{z}_i\right)$ respectively. The parameters h_1^b and h_1^o depict the dependence of the unobserved heterogeneities on the initial conditions.

Substituting (4.16) into (4.1) leads to a final underlying latent variable specification:

$$y_{it}^{b*} = \phi_1 y_{it-1}^b + \gamma_1' x_{it} + \beta_1' w_i + h_1^b y_{i0}^b + h_0^b + h_2^b \bar{x}_i + h_3^b x_{i0} + \delta_{1i} + e_{it}^b \tag{4.20}$$

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

Substituting (4.17) into (4.7) leads to a final underlying latent variable specification:

$$y_{it}^{o*} = \phi_2 y_{it-1}^o + \gamma_2' z_{it} + \beta_2' v_i + h_1^o y_{i0}^0 + h_0^o + h_2^o \bar{z}_i + h_3^o z_{i0} + \delta_{2i} + e_{it}^o \tag{4.21}$$

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

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

$$\Sigma_{\varepsilon b \varepsilon o} = \begin{pmatrix} \sigma_{eb}^2 & \rho_{ebeo}\sigma_{eb}\sigma_{eo} \\ \rho_{ebeo}\sigma_{eb}\sigma_{eo} & \sigma_{eo}^2 \end{pmatrix} \tag{4.22}$$

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

$$\Sigma_{\delta_{1i}\delta_{2i}} = \begin{pmatrix} \sigma_1^2 & \rho_{\delta_{1i}\delta_{2i}}\sigma_1\sigma_2 \\ \rho_{\delta_{1i}\delta_{2i}}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \tag{4.23}$$

The entries of the matrices are also estimated.

The dummies are removed from x_i and z_i to elude perfect collinearity. The estimates of h_1^b and h_1^o indicate the association between the unobserved heterogeneities and first observation.

The response probabilities of each value of y_{it} are hence:

$$P(y_{it}=k) = \begin{cases} P(y_{it}=0) \\ = \left(1-\Phi\left(\phi_1 y_{it-1}^b + \gamma'_1 x_{it} + \beta'_1 w_i + h_0^b + h_1^b y_{i0}^b + h_2^b \bar{x}_i + h_3^b x_{i0} + \delta_{1i}\right)\right) \\ +\Phi\left(\phi_1 y_{it-1}^b + \gamma'_1 x_{it} + \beta'_1 w_i + h_1^b y_{i0}^{b*} + h_0^b + h_2^b \bar{x}_i + h_3^b x_{i0} + \delta_{1i},\right. \\ \left. -\phi_2 y_{it-1}^o - \gamma'_2 z'_{it} - \beta'_2 v_i - h_1^o y_{i0}^{0*} + h_0^o + h_2^o \bar{z}_i + h_3^o z_{i0} + \delta_{2i}, -\rho_{ebeo}\right) \\ \text{if } k=0, \\ P(y_{it}=k|x,\, z) = \\ \left[\Phi\left(\phi_1 y_{it-1}^b + \gamma'_1 x_{it} + \beta'_1 w_i + h_0^b + h_1^b y_{i0}^b + h_2^b \bar{x}_{ii} + h_3^b x_{i0} + \delta_{1i},\right.\right. \\ \left.\tau_k - \phi_2 y_{it-1}^o - \gamma'_2 z_{it} - \beta'_2 v_i - h_0^o - h_1^o y_{i0}^0 - h_2^o \bar{z}_i - h_3^o z_{i0} \delta_{2i}, -\rho_{ebeo}\right) - \\ \Phi\left(\phi_1 y_{it-1}^b + \gamma'_1 x_{it} + \beta'_1 w_i + h_0^b + h_1^b y_{i0}^b + h_2^b \bar{x}_i + h_3^o z_{i0} \delta_{1i},\right. \\ \left.\left.\tau_{k-1} - \phi_2 y_{it-1}^o - \gamma'_2 z_{it} - \beta'_2 v_i - h_0^o - h_1^o y_{i0}^0 - h_2^o \bar{z}_i - h_3^o z_{i0} \delta_{2i}, -\rho_{ebeo}\right)\right] \\ \text{if k} = 1,\ 2,\ \ldots,\ \text{K}-1, \\ P(y_{it}=K|x,\, z) = \\ \left[\Phi\left(\phi_1 y_{it-1}^b + \gamma'_1 x_{it} + \beta'_1 w_i + h_0^b + h_1^b y_{i0}^b + h_2^b \bar{x}_i + h_3^b x_{i0} \delta_{1i},\right.\right. \\ \left.\left.\phi_2 y_{it-1}^o + \gamma'_2 z_{it} + \beta'_2 v_i + h_0^o + h_1^o y_{i0}^0 + h_2^o \bar{z}_i + h_3^o z_{i0} \delta_{2i} - \tau_K, \rho_{ebeo}\right)\right] \\ \text{if } k=K \end{cases} \tag{4.24}$$

Independence is an important assumption in the derivation of an ordinary probit model, that is, the joint probability for the data equals the product of the marginal probabilities. The log-likelihood function is the sum of the specific log-likelihood contributions. However, this does not apply to serially dependent data since y_{it}^* is a function of y_{it-1} and $P(y_{it}=k)$ is correlated with $P(y_{it-r}=k)$. Therefore, joint probability of the data and log-likelihood are not the product of the time-specific probabilities and sum of the time-specific log-likelihood contributions respectively.

Wooldridge (2005) indicate that the assumption that $y_{it} = \left(y_{it}^{b},\ y_{it}^{o}\right)$ relies on its once lagged value y_{it-1} but not on any of its other lags implies that the joint distribution of $y_{i1},\ y_{i2},\ ...,\ y_{iT}$ conditional on y_{i0}, $\kappa_i = (\kappa_{1i},\ \kappa_{2i})$, $(x_{it},\ \mathrm{z}_{it})$ and $(w_i,\ v_i)$ is denoted as

$$\prod_{i=1}^{N}\prod_{t=1}^{T} f\left(y_{it}|y_{it-1},\ x_{it},\ z_{it}, w_i, v_i,\ \kappa_i\right) \tag{4.25}$$

Conditioned on the unobserved heterogeneities $\kappa_i = (\kappa_{1i},\ \kappa_{2i})$, the response on y_{it} are considered to be uncorrelated.

The parameter to be estimated in this model will be denoted by

$$\Theta = \left(\phi_1,\ \phi_2,\ \gamma_1,\ \gamma_2, \beta_1,\ \beta_2, \rho_{ebeo}, \rho_{\delta_{1i}\delta_{2i}},\ \tau,\ h_0^b,\ h_1^b,\ h_2^b,\ h_0^o,\ h_1^o, h_2^o,\ \sigma_1,\ \sigma_2\right) \tag{4.26}$$

The likelihood function of respondent i, beginning from $t = 1$ and conditional on the covariates and the initial conditions, is denoted as,

$$L_i = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\prod_{t=1}^{T}\prod_{k=0}^{K} L_{itk}\left(y_{itk}^{b},\ y_{itk}^{o}|y_{i0k}^{b},\ y_{ikt-1}^{b}, x_{it}, \bar{x}_i,\ w_i,\right.$$
$$\left. y_{i0k}^{o},\ y_{ikt-1}^{o},\ z_{it}, \bar{z}_i,\ v_i,\ \delta_{1i},\ \delta_{2i}\right) g\left(\delta_{1i}, \delta_{2i}\right) d\delta_{1i} d\delta_{2i} \tag{4.27}$$

Where $\prod_{t=1}^{T}\prod_{k=0}^{K} L_{itk}\left(y_{itk}^{b},\ y_{itk}^{o}|y_{i0k}^{b},\ y_{ikt-1}^{b},\ x_i,\ w_i,\ y_{i0k}^{o},\ y_{ikt-1}^{o},\ z_i,\ v_i,\ \delta_{1i},\ \delta_{2i}\right)$ and $g\left(\delta_{1i},\ \delta_{2i}\right)$ denote respectively the likelihood function of respondent i conditional on the unobserved heterogeneities, and the bivariate normal density function of $\left(\delta_{1i},\ \delta_{2i}\right)$ Assuming a normal distribution for these individual random effects with variances $\sigma_{\delta_1}^2$ and $\sigma_{\delta_2}^2$ respectively and a correlation coefficient $\rho_{\delta_{1i}\delta_{2i}}$, the density function for the individual effects is given by:

$$g\left(\delta_{1i},\delta_{2i}|\sigma_1^2,\sigma_2^2,\rho_{\delta_{1i}\delta_{2i}}\right) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}}\exp\left\{\frac{-1}{2\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)} * \left[\left(\frac{\delta_{1i}}{\sigma_1}\right)^2 - 2\rho_{\delta_{1i}\delta_{2i}}\left(\frac{\delta_{1i}}{\sigma_1}\right)\left(\frac{\delta_{2i}}{\sigma_2}\right) + \left(\frac{\delta_{2i}}{\sigma_2}\right)^2\right]\right. \tag{4.28}$$

This density function does not depend on observables but on the three parameters which should be estimated.

The unconditional (to the individual random effects) joint density for the i^{th} individual is obtained by averaging over the distribution of these unobserved heterogeneities:

$$L_i = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\prod_{t=1}^{T}\prod_{k=0}^{K} L_{tk}\left(y_{tk}^b,\ y_{tk}^o|y_{i0k}^b,\ y_{ikt-1}^b,\ x_{it},\ \bar{x}_i,\ w_i,\ y_{i0k}^o,\ y_{ikt-1}^o, z_{it},\ \bar{z}_i,\ v_i,\ \delta_{1i},\ \delta_{2i}\right)\frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}}\exp \left\{\frac{-1}{2\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}\left[\left(\frac{\delta_1}{\sigma_1}\right)^2 - 2\rho_{\delta_{1i}\delta_{2i}}\left(\frac{\delta_{1i}}{\sigma_1}\right)\left(\frac{\delta_{2i}}{\sigma_2}\right) + \left(\frac{\delta_{2i}}{\sigma_2}\right)^2\right\} d\delta_{1i}d\delta_{2i} \tag{4.29}$$

or

$$L_i = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\prod_{t=1}^{T}\prod_{k=0}^{K}\left\{P\left(y_{it}=k|y_{i0k}^b,\ y_{ikt-1}^b, y_{i0k}^o,\ y_{ikt-1}^o, x_{it}, z_{it},\ \bar{x}_i, \overline{z}_i, w_i, v_i, \Theta\right)\right\}^{d_{itk}} \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}}\exp\left\{\frac{-1}{2\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}\left[\left(\frac{\delta_1}{\sigma_1}\right)^2 - 2\rho_{\delta_{1i}\delta_{2i}}\left(\frac{\delta_{1i}}{\sigma_1}\right)\left(\frac{\delta_{2i}}{\sigma_2}\right) + \left(\frac{\delta_{2i}}{\sigma_2}\right)^2\right\} d\delta_{1i}d\delta_{2i} \tag{4.30}$$

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

4.3 Likelihood Approximation by Gauss–Hermite quadrature

Consider the binary and ordered probit models given by equations (4.20) and (4.21). Suppose the vectors $\left(e_{it}^{b},\ e_{it}^{o}\right)$ and $(\delta_{1i},\ \delta_{2i})$ are uncorrelated with each other and independently and identically distributed over time and across respondents assuming a bivariate normal distributions with mean zero and covariance matrix given by $\Sigma_{\varepsilon b\varepsilon o}$ and $\Sigma_{\delta_{1i}\delta_{2i}}$ respectively. The likelihood function of respondent i, beginning from $t=1$ and conditional on the covariates and the initial conditions, is obtained by "integrating out" the unobserved heterogeneities. The likelihood function of respondent i conditional on the unobserved heterogeneities is given in equation (3.30) with σ_1^2 and σ_2^2 being normalized to one. The bivariate normal distribution of δ_{1i} and δ_{2i} is given by equation (4.28)

Hence, equation (4.30) can be written as,

$$L_i = \int_{-\infty}^{+\infty} \exp\left\{\frac{-1}{2\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}\frac{\delta_{2i}^2}{\sigma_2^2}\right\}\prod_{t=1}^{T}\prod_{k=0}^{K} M\left(\delta_{1i}\right) d\delta_{2i} \tag{4.31}$$

Where

$$\begin{aligned} M\left(\delta_{1i}\right) = & \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}} \\ | & \int_{-\infty}^{+\infty} \exp\left\{\frac{-1}{2\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}\frac{\delta_{1i}^2}{\sigma_1^2}\right\} \exp\left\{\frac{1}{\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}\rho_{\delta_1\delta_2}\left(\frac{\delta_{1i}}{\sigma_1}\right)\left(\frac{\delta_{2i}}{\sigma_2}\right)\right\} \\ & \prod_{t=1}^{T}\prod_{k=0}^{K}\left\{P\left(y_{it}=k|y_{i0k}^{b},\ y_{ikt-1}^{b}, y_{i0k}^{o},\ y_{ikt-1}^{o}, x_{it}, z_{it},\ \overline{x}_i, \overline{z}_i, w_i, v_i, \theta\right)\right\}^{d_{itk}} d\delta_{1i} \end{aligned} \tag{4.32}$$

Equation (4.32) can be approximated by "two-step" Gauss-Hermite quadrature that is given by

$$\int_{-\infty}^{+\infty} e^{-z^2} f(z)\, dz \simeq \sum_{h=1}^{H} w_h f\left(a_h\right) \tag{4.33}$$

where $a_h$and w_h are nodes and weights respectively. The values of nodes and weights are found in the table given in the mathematical textbooks by Abramovitz and Stegun (1964) , and H is the actual number of nodes and weights. The accuracy depend on the value of H. However, a large H raises the cost of computation significantly.

The procedure involves, in the first step, in approximating equation (4.32) using equation (4.33). In the succeeding step, a second approximation is applied to equation (4.32) where $M(\delta_{1i})$ is replaced by its first-step Gauss-Hermite approximation.

Let A and B denote by $A = \phi_1 y^{b*}_{it-1} + \gamma'_1 x_{it} + \beta'_1 w_i + h^b_0 + h^b_1 y^{b*}_{i0} + h^b_2 \overline{x}_i$ and $B = \phi_2 y^{o*}_{it-1} + \gamma'_2 z_{it} + \beta'_2 v_i + h^o_0 + h^o_1 y^{0*}_{i0} + h^o_2 \overline{z}_i$ respectively. Then equation (4.24) becomes,

$$P(y_{it} = k) = \begin{cases} P(y_{it} = 0) = (1 - \Phi(A + \delta_{1i})) \\ +\Phi(A + \delta_{1i},\ -B - \delta_{2i},\ -\rho_{ebeo}) \\ k = 0, \\ P(y_{it} = k|x,\ z) = \\ \Phi(A + \delta_{1i},\ \tau_k - B - \delta_{2i},\ -\rho_{ebeo}) \\ -\Phi(A + \delta_{1i},\ \tau_{k-1} - B_i - \delta_{2i},\ -\rho_{ebeo}) \\ \text{if k} = 1,\ 2,\ ...,\ \text{K} - 1, \\ P(y_{it} = K|x,\ z) = \\ \Phi(A + \delta_{1i},\ B + \delta_{2i} - \tau_{K-1},\ \rho_{ebeo}) \\ \text{if } k = K \end{cases} \tag{4.34}$$

Then,

$$M(\delta_{1i}) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2\left(1 - \rho^2_{\delta_{1i}\delta_{2i}}\right)}} \int_{-\infty}^{+\infty} \exp\left\{\frac{-\delta^2_{1i}}{2\sigma_1^2\left(1 - \rho^2_{\delta_{1i}\delta_{2i}}\right)}\right\} \exp\left\{\frac{1}{\left(1 - \rho^2_{\delta_{1i}\delta_{2i}}\right)}\rho_{\delta_{1i}\delta_{2i}}\left(\frac{\delta_{1i}}{\sigma_1}\right)\left(\frac{\delta_{2i}}{\sigma_2}\right)\right\} \prod_{k=0}^{K}\left\{P\left(y_{it} = k|y^{*b}_{i0k},\ y^{*b}_{ikt-1}, y^{*o}_{i0k},\ y^{*o}_{ikt-1}, x_{it}, z_{it},\ \overline{x}_i, \overline{z}_i, w_i, v_i, \theta\right)\right\}^{d_{itk}} d\delta_{1i} \tag{4.35}$$

Consider the first change of variable $z_{1i} = \frac{\delta_{1i}}{\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}}$ and $\delta_{1i} = z_{1i}\sigma_1\sqrt{2\left(1-\rho^2_{\delta_1\delta_2}\right)}$

. Differentiate z_{1i} with respect to δ_{1i}, we get,

$\frac{dz_{1i}}{d\delta_{1i}} = \frac{1}{\sigma_1\sqrt{2\left(1-\rho^2_{\delta_1\delta_2}\right)}}$ or $\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}dz_{1i} = d\delta_{1i}$

Substituting it in equation (4.35) we get,

$$M(\delta_{2i}) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}}\int_{-\infty}^{+\infty}\exp\{-z_{1i}\}^2\exp$$
$$\left\{\frac{\rho_{\delta_{1i}\delta_{2i}}}{\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\left(\frac{z_{1i}\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}}{\sigma_1}\right)\left(\frac{\delta_{2i}}{\sigma_2}\right)\right\}\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\sigma_1 dz_{1i}$$
$$\prod_{k=0}^{K}\left\{P\left(y_{it}=k|y^{*b}_{i0k},\ y^{*b}_{ikt-1},y^{*o}_{i0k},\ y^{*o}_{ikt-1},x_{it},z_{it},\ \bar{x}_i,\ \overline{z}_i,w_i,v_i,\theta\right)\right\}^{d_{itk}} \tag{4.36}$$

Simplifying the above equation we get,

$$M(\delta_{2i}) = \frac{\sqrt{2}}{2\pi\sigma_{\delta_2}}\int_{-\infty}^{+\infty}\exp\left\{-z^2_{1i}\right\}\exp\left\{\frac{\rho_{\delta_{1i}\delta_{2i}}}{\left(1-\rho^2_{\delta_{1i}\delta_i}\right)}z_{1i}\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\left(\frac{\delta_{2i}}{\sigma_2}\right)\right\}$$
$$\prod_{k=0}^{K}\left\{\left(1-\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\right)\right)+\right.$$
$$+\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},-B-\delta_{2i},-\rho_{ebeo}\right)$$
$$\Phi\left(A+a_h\sigma_{\delta_1}\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ \tau_k-B-\delta_{2i},-\rho_{ebeo}\right)$$
$$-\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ \tau_{k-1}-B-\delta_{2i},-\rho_{ebeo}\right)$$
$$\left.\left[\Phi\left(A+z_{1i}\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ B+\delta_{2i}-\tau_{K-1},\ \rho_{ebeo}\right)\right]\right\}^{d_{itk}} dz_{1i} \tag{4.37}$$

and can be approximated using equation (4.33) by

$$M(\delta_{2i}) = \frac{\sqrt{2}}{2\pi\sigma_2}\sum_{h=1}^{H} w_h \exp\left\{\frac{\rho_{\delta_{1i}\delta_{2i}}}{\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)} a_h\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\left(\frac{\delta_{2i}}{\sigma_2}\right)\right\}$$
$$\prod_{k=0}^{K}\left\{\left(1-\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\right)\right)+\right.$$
$$+\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}, -B-\delta_{2i}, -\rho_{ebeo}\right)$$
$$\Phi\left(A+a_h\sigma_{\delta_1}\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ \tau_k-B-\delta_{2i}, -\rho_{ebeo}\right)$$
$$-\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ \tau_{k-1}-B-\delta_{2i}, -\rho_{ebeo}\right)$$
$$\left.\left[\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ B+\delta_{2i}-\tau_{K-1},\ \rho_{ebeo}\right)\right]\right\}^{d_{itk}} \tag{4.38}$$

where a_h and w_h are the nodes and weights respectively. Consider the second change of variable, $z_{2i} = \frac{\delta_{2i}}{\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}}$ and $\delta_{2i} = z_{2i}\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}$

Differentiate z_{2i} with respect to δ_{2i}, we get, $\frac{dz_{2i}}{d\delta_{2i}} = \frac{1}{\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}}$ and

$\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}dz_{2i} = d\delta_{2i}$

Replacing the derivative in equation (4.38), we get,

$$L_i = \int_{-\infty}^{+\infty}\exp\left\{\left(\frac{-\delta_{2i}}{\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}}\right)^2\right\}\prod_{t=1}^{T}M(\delta_{1i})\left(\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\right)dz_{2i} \tag{4.39}$$

we get,

$$L_i = \int_{-\infty}^{+\infty}\exp\left\{-{z_{2i}}^2\right\}\prod_{t=1}^{T}M(\delta_{1i})\left(\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\right)dz_{2i} \tag{4.40}$$

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

$$L_i = \int_{-\infty}^{+\infty} \exp\left\{-{z_{2i}}^2\right\} \frac{\sqrt{2}}{2\pi\sigma_2} \sum_{h=1}^{H} w_h \exp$$
$$\left\{ \frac{\rho_{\delta_{1i}\delta_{2i}}}{\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)} a_h \sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)} \left(\frac{z_{2i}\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}}{\sigma_2} \right) \right\}$$
$$\left(\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\right) dz_{2i} \prod_{t=1}^{T}\prod_{k=0}^{K} \left\{ \left[\left(1-\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\right)\right)\right.\right.$$
$$\left.+\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}, -B-\delta_{2i}, -\rho_{ebeo}\right)\right]$$
$$\left[\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ \tau_k-B-\delta_{2i}, -\rho_{ebeo}\right)-\right.$$
$$\left.\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ \tau_{k-1}-B_i-\delta_{2i}, -\rho_{ebeo}\right)\right]$$
$$\left.\left.\left[\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ B+\delta_{2i}-\tau_{K-1},\ \rho_{ebeo}\right)\right]\right\}\right\}^{d_{itk}} \tag{4.41}$$

Simplifying equation (4.41), we get

$$L_i = \int_{-\infty}^{+\infty} \exp\left\{-{z_{2i}}^2\right\} \prod_{t=1}^{T} \pi^{-1}\sqrt{\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)} \sum_{h=1}^{M} w_h \exp\left\{2\rho_{\delta_{1i}\delta_{2i}} z_{2i} a_h\right\}$$
$$\prod_{k=0}^{K}\left\{\left[\left(1-\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\right)\right)\right.\right.$$
$$\left.+\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}, -B-z_{2i}\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}, -\rho_{ebeo}\right)\right]$$
$$\left[\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ \tau_k-B-z_{2i}\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}, -\rho_{ebeo}\right)-\right.$$
$$\left.\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ \tau_{k-1}-B_i-z_{2i}\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}, -\rho_{ebeo}\right)\right]$$
$$\left[\Phi\left(A+a_h\sigma_{\delta_1}\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ B+z_{2i}\sigma_2\sqrt{2\left(1-\rho^2_{\delta_1\delta_2}\right)}-\tau_{K-1},\ \rho_{ebeo}\right)\right]$$
$$\left.\right\}^{d_{itk}} dz_{2i} \tag{4.42}$$

that can be approximated by equation (4.33)

$$L_i = \pi^{-1}\sqrt{\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\sum_{q=1}^{Q} w_q \sum_{h=1}^{H}\left\{w_h exp\left\{2\rho_{\delta_{1i}\delta_{2i}} a_q a_h\right\}\right.$$
$$\prod_{t=1}^{T}\prod_{k=0}^{K}\left\{\left[\left(1-\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\right)\right)\right.\right.$$
$$\left.+\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}, -B-a_q\sigma 2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}, -\rho_{ebeo}\right)\right]$$
$$\left[\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ \tau_k-B-a_q\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}, -\rho_{ebeo}-\right)\right.$$
$$\left.\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ \tau_{k-1}-B_i-a_q\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}, -\rho_{ebeo}\right)\right]$$
$$\left[\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ B+a_q\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}-\tau_{K-1},\ \rho_{ebeo}\right)\right]$$
$$\}\}^{d_{itk}} \tag{4.43}$$

or

$$L_i = \pi^{-1}\sqrt{\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\sum_{q=1}^{Q} w_q \sum_{h=1}^{H}\left\{w_h exp\left\{2\rho_{\delta_{1i}\delta_{2i}} a_q a_h\right\}\right.$$
$$\prod_{t=1}^{T}\prod_{k=0}^{K}\left\{\left[\left(1-\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\right)\right)\right.\right.$$
$$\left.+\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}, -B-a_q\sigma 2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}, -\rho_{ebeo}\right)\right]$$
$$\left[\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ \tau_k-B-a_q\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}, -\rho_{ebeo}\right)-\right.$$
$$\left.\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ \tau_{k-1}-B_i-a_q\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}, -\rho_{ebeo}\right)\right]$$
$$\left[\Phi\left(A+a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ B+a_q\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}-\tau_{K-1},\ \rho_{ebeo}\right)\right]$$
$$\}\}^{d_{itk}} \tag{4.44}$$

where a_q and w_q are the nodes and weights respectively.

The product over i of the approximate likelihood function can be maximized using Newton-Raphson Method procedures to find the estimates of the parameters.

Finally, as the respondents are uncorrelated, the log-likelihood function for N individuals should be expressed as:

$$
\begin{aligned}
l(\Theta, y) = \sum_{i=1}^{N} & \left\{ \log \left\{ \pi^{-1} \sqrt{\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)} \sum_{q=1}^{Q} \sum_{h=1}^{H} w_q w_h exp\left\{2\rho_{\delta_{1i}\delta_{2i}} a_q a_h\right\} \right. \right. \\
& \prod_{t=1}^{T} \prod_{k=0}^{K} \left\{ \left[\left(1-\Phi\left(A + a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\right)\right) \right. \right. \\
& \left. +\Phi\left(A + a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}, -B - a_q\sigma 2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}, -\rho_{ebeo}\right)\right] \\
& \left[\Phi\left(A + a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ \tau_k - B - a_q\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}, -\rho_{ebeo}\right) - \right. \\
& \left. \Phi\left(A + a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ \tau_{k-1} - B_i - a_q\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}, -\rho_{ebeo}\right)\right] \\
& \left[\Phi\left(A + a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)},\ B + a_q\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)} - \tau_{K-1},\ \rho_{ebeo}\right)\right] \\
& \}\}^{d_{itk}}
\end{aligned}
\tag{4.45}
$$

Let

$$u_{1,qh} = \phi_1 y^b_{it-1} + \gamma_1' x_{it} + \beta_1 w_i + h^b_0 + h^b_1 y^b_{i0} + h^b_2 x_i + h^b_3 x_{i0} + a_h\sigma_1\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}$$

$$u_{1,qh0} = \tau_0 - \phi_2 y^o_{it-1} - \gamma_2' z_{it} - \beta_2 v_i - h^o_0 - h^o_1 y^0_{i0} - h^o_2 z_i - h^o_3 z_{i0} - a_q\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}$$

$$u_{1,qhk} = \tau_k - \phi_2 y^o_{it-1} - \gamma_2' z_{it} - \beta_2 v_i - h^o_0 - h^o_1 y^0_{i0} - h^o_2 z_i - h^o_3 z_{i0} - a_q\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}$$

$$u_{1,qhk-1} = \tau_{k-1} - \phi_2 y^o_{it-1} - \gamma_2' z_{it} - \beta_2 v_i - h^o_0 - h^o_1 y^0_{i0} - h^o_2 z_i - h^o_3 z_{i0} - a_q\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}$$

$$u_{1,qhK} = \phi_2 y^o_{it-1} + \gamma_2' z_{it} + \beta_2 v_i + h^o_0 + h^o_1 y^0_{i0} + h^o_2 z_i + h^o_3 z_{i0} + a_q\sigma_2\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)} - \tau_{K-1}$$

The study present the auxiliary conditional distribution of κ_{1i} with a constant h^b_0. Therefore, the study removed the constant in the structural equation. Similarly, the study present the auxiliary conditional distribution of κ_{2i} with a constant h^o_0. Therefore, the study removed the constant in the structural equation.

Using $u_{1,qh}$, $u_{1,qh0}$, $u_{1,qhk}$, $u_{1,qhk-1}$ and $u_{1,qhK}$ equations in (4.45), we get,

$$\begin{aligned}\ell(\Theta,y) = \sum_{i=1}^{N}\log\Bigg\{ & \pi^{-1}\sqrt{\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\sum_{q=1}^{Q}\sum_{h=1}^{M} w_q w_h exp\left\{2\rho_{\delta_1\delta_2}a_q a_h\right\} \\ & \prod_{t=1}^{T}\prod_{k=0}^{K}\left\{\left[\left(\left(1-\Phi\left(u_{1,qh}\right)\right)+\Phi\left(u_{1,qh},\ u_{1,qh0},-\rho_{ebeo}\right)\right)\right]\right. \\ & \left[\Phi\left(u_{1,qh},\ u_{1,qhk},-\rho_{ebeo}\right)-\Phi\left(u_{1,qh},\ u_{1,qhk-1},-\rho_{ebeo}\right)\right] \\ & \left.\left[\Phi\left(u_{1,qh},\ u_{1,qK},\rho_{ebeo}\right)\right]\right\}\Bigg\}^{d_{itk}}\end{aligned} \tag{4.46}$$

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

$$\begin{aligned}\ell(\Theta,y) = \sum_{i=1}^{N}\log\Bigg\{ & \pi^{-1}\sqrt{\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\sum_{q=1}^{Q}\sum_{h=1}^{H} w_q w_h exp\left\{2\rho_{\delta_1\delta_2}a_q a_h\right\} \\ & \prod_{t=1}^{T}\prod_{k=0}^{K}\left\{\left[\left(\left(1-\Phi\left(u_{1,qh}\right)\right)+\Phi\left(u_{1,qh}\right)\Phi\left(u_{1,qh0}\right)\right)\right]\right. \\ & \left.\left[\Phi\left(u_{1,qh}\right)\left(\Phi\left(u_{1,qhk}\right)-\Phi\left(u_{1,qhk-1}\right)\right)\right]\ \left[\Phi\left(u_{1,qh}\right)\Phi\left(u_{1,qK}\right)\right]\right\}\Bigg\}^{d_{itk}}\end{aligned} \tag{4.47}$$

To evaluate the likelihood function at each point, 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 apply Newton-Raphson. Newton-Raphson method is a gradient based root finding methods that may be used to determine extreme points of a function, that is, optimization. The loglikelihood in many problems approaches a quadratic function spreading at its maximum. Newton-Raphson method possess quadratic convergence characteristics. Therefore, convergence is fast. The number of iterations is independent of the size of the system. The solutions to a high accuracy is obtained nearly always in few iterations. Overall, there is saving time in computation time since fewer numbers of iterations are required. The Newton-Raphson method approximates these MLE's by using the following procedure:

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

where $t = 0,\ 1,\ 2,\ ...$ is the iteration number. To apply the Newton-Raphson algorithm we require to calculate the Hessian matrix at every iteration which consists of the following expressions

$$\begin{pmatrix} -\frac{\partial^2 \ell(\psi,\lambda,y)}{\partial\psi\partial\psi^t} & -\frac{\partial^2 \ell(\psi,\lambda,y)}{\partial\psi\partial\lambda} \\ -\frac{\partial^2 \ell(\psi,\lambda,y)}{\partial\psi\partial\lambda} & -\frac{\partial^2 \ell(\psi,\lambda,y)}{\partial\lambda\partial\lambda^t} \end{pmatrix}$$

The log-likelihood function can be maximised by applying

$$\Theta^{t+1} = \Theta^t - H\left(\Theta^t\right)^{-1} G\left(\Theta^t\right) \tag{4.48}$$

until convergence is attained, where $G(\Theta^t)$ is the gradient function computed at Θ^t and is the Hessian matrix computed at $H(\Theta^t)$. This implies that starting values Θ^0 are required. In the subsequent subsections, the entries of Hessian matrix from the log-likelihood with respect to Θ are obtained. The algorithm is assumed to have converged when the log likelihood changes by a small constant $\varepsilon > 0$, that is, $\left|\Theta^{t+1} - \Theta^t\right| < \varepsilon$ where, for instance, $\varepsilon = 10^{-5}$. According to Efron and Hinkley (1978), the standard errors of the estimates are computed as the square root of the observed information matrix, i.e. negative of the second-order differentiation of the log-likelihood function. The estimates are asymptotically normally distributed:

$$\hat{\Theta}_{ML} \sim N\left(\Theta_0,\ H^{-1}\right) \tag{4.49}$$

4.4 First -Order Differentiations of the Log-Likelihood Function

Let

$$f\left(y_{itk}|\Theta\right)=\pi^{-1}\sqrt{\left(1-\rho^{2}_{\delta_{1i}\delta_{2i}}\right)}exp\left\{2\rho_{\delta_{1i}\delta_{2i}}a_{q}a_{h}\right\}$$
$$\prod_{t=1}^{T}\prod_{k=0}^{K}\left\{\left[\left(\left(1-\Phi\left(u_{1,qh}\right)\right)+\Phi\left(u_{1,qh},\ u_{1,qh0},-\rho_{ebeo}\right)\right)\right]\right.$$
$$\left[\Phi\left(u_{1,qh},\ u_{1,qhk},-\rho_{ebeo}\right)-\Phi\left(u_{1,qh},\ u_{1,qhk-1},-\rho_{ebeo}\right)\right]$$
$$\left.\left[\Phi\left(u_{1,qh},\ u_{1,qK}-\tau_{K-1},\rho_{ebeo}\right)\right]\right\}^{d_{itk}}\Big\} \tag{4.50}$$

then the first derivative with respect to parameter $\gamma_1 \in \Theta$ of equation (4.46) is,

$$\frac{\partial\ell\left(\Theta,y\right)}{\partial\gamma_{1}}=\frac{\partial}{\partial\gamma_{1}}\left(\sum_{i=1}^{N}\log\sum_{q=1}^{Q}\sum_{h=1}^{H}w_{q}w_{h}f\left(y_{itk}|\Theta\right)\right)=\sum_{i=1}^{N}\frac{\sum\limits_{q=1}^{Q}\sum\limits_{h=1}^{H}w_{q}w_{h}\frac{\partial f\left(y_{itk}|\Theta\right)}{\partial\gamma_{1}}}{\sum\limits_{q=1}^{Q}\sum\limits_{h=1}^{H}w_{q}w_{h}f\left(y_{itk}|\Theta\right)} \tag{4.51}$$

Using the logarithm, the equation can also be writtern as,

$$\frac{\partial\ell\left(\Theta,y\right)}{\partial\gamma_{1}}=\sum_{i=1}^{N}\frac{\sum\limits_{q=1}^{Q}\sum\limits_{h=1}^{H}w_{q}w_{h}f\left(y_{itk}|\Theta\right)\frac{\partial\log f\left(y_{itk}|\Theta\right)}{\partial\gamma_{1}}}{\sum\limits_{r=1}^{Q}\sum\limits_{l=1}^{H}w_{q}w_{h}f\left(y_{itk}|\Theta\right)}=\sum_{i=1}^{N}\sum_{q=1}^{Q}\sum_{h=1}^{H}\pi_{qhi}\frac{\partial\log f\left(y_{itk}|\Theta\right)}{\partial\gamma_{1}} \tag{4.52}$$

where $\pi_{qhi}=\frac{w_{q}w_{h}f\left(y_{itk}|\Theta\right)}{\sum\limits_{q=1}^{Q}\sum\limits_{h=1}^{H}w_{q}w_{h}f\left(y_{itk}|\Theta\right)}$.

Let $\Phi\left(A,B\right)$ be a bivariate cumulative distribution function where A and B are standard normal random variables with correlation ρ . Then, the conditional probability density function of Y given Z = z is given by

$$\frac{\partial\Phi\left(A,B,\ \rho\right)}{\partial A}=\frac{\partial}{\partial A}\int_{-\infty}^{A}\int_{-\infty}^{B}\phi\left(s,t,\rho\right)dsdt=\int_{-\infty}^{B}\phi\left(A,t,\rho\right)dt=\phi\left(A\right)\Phi\left(\frac{B-\rho A}{\sqrt{1-\rho^{2}}}\right) \tag{4.53}$$

$$\frac{\partial \Phi(A,\ B,\ \rho)}{\partial \lambda} = \frac{\partial \Phi(A, B,\ \rho)}{\partial A}\frac{\partial A}{\partial \lambda} = \phi(A)\,\Phi\left(\frac{B-\rho A}{\sqrt{1-\rho^2}}\right)\frac{\partial A}{\partial \lambda} \tag{4.54}$$

Let the log of $f(y_{itk}|\Theta)$ be given by,

$$\begin{aligned}\log f(y_{itk}|\Theta) = &-\log \pi + \frac{1}{2}\log\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right) + 2\rho_{\delta_{1i}\delta_{2i}} a_q a_h + \\ &\sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk} \log\left\{\left[\left(\left(1-\Phi\left(u_{1,qh}\right)\right) + \Phi\left(u_{1,qh}\right)\Phi\left(u_{1,qh0}\right)\right)\right]\right. \\ &\left[\Phi\left(u_{1,qh},\ u_{1,qhk}, -\rho_{ebeo}\right) - \Phi\left(u_{1,qh},\ u_{1,qhk-1}, -\rho_{ebeo}\right)\right] \\ &\left.\left[\Phi\left(u_{1,qh},\ u_{1,qK}, \rho_{ebeo}\right)\right]\right\}\end{aligned} \tag{4.55}$$

In case the errors terms are independent, we get

$$\begin{aligned}\log f(y_{itk}|\Theta) = &-\log \pi + \frac{1}{2}\log\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right) + 2\rho_{\delta_{1i}\delta_{2i}} a_q a_h + \\ &\sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk} \log\left\{\left[\left(\left(1-\Phi\left(u_{1,qh}\right)\right) + \Phi\left(u_{1,qh}\right)\Phi\left(u_{1,qh0}\right)\right)\right]\right. \\ &\left.\left[\Phi\left(u_{1,qh}\right)\left(\Phi\left(u_{1,qhk}\right) - \Phi\left(u_{1,qhk-1}\right)\right)\right]\left[\Phi\left(u_{1,qh}\right)\Phi\left(u_{1,qK}\right)\right]\right\}\end{aligned} \tag{4.56}$$

Let

$$\begin{aligned}C = &\left(\left(1-\Phi\left(u_{1,qh}\right)\right) + \Phi\left(u_{1,qh},\ u_{1,qh0}, -\rho_{ebeo}\right)\right) \\ &\left(\Phi\left(u_{1,qh},\ u_{1,qhk}, -\rho_{ebeo}\right) - \Phi\left(u_{1,qh},\ u_{1,qhk-1}, -\rho_{ebeo}\right)\right)\ \left(\Phi\left(u_{1,qh},\ u_{1,qK},\ \rho_{ebeo}\right)\right)\end{aligned} \tag{4.57}$$

Let

$$\theta = \phi_1,\ \gamma_1,\ \beta_1,\ h_0^b,\ h_1^b,\ h_2^b$$

$$\psi = \phi_2,\ \gamma_2,\ \beta_2,\ h_0^o,\ h_1^o,\ h_2^o$$

$$X = y_{it-1}^b,\ x_{it},\ w_i,\ 1,\ y_{i0}^b,\ x_i$$

$$Z = \ -y_{it-1}^o,\ -z_{it},\ -v_i,\ -1,\ -y_{i0}^o,\ -z_i$$

Let φ denote a probability density function.

The differentiation of (4.55) with respect to θ is given by,

$$\begin{aligned}\frac{\partial \log f(y_{itk}|\Theta)}{\partial \theta} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\Bigg\{ & \left[-\varphi\left(u_{1,qh}\right)+\varphi\left(u_{1,qh}\right)\Phi\left(\frac{u_{1,qh0}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^{2}}}\right)\right] \\ & +\left[\varphi\left(u_{1,qh}\right)\Phi\left(\frac{u_{1,qhk}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^{2}}}\right)-\varphi\left(u_{1,qh}\right)* \right. \\ & \left. \Phi\left(\frac{u_{1,qhk-1}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^{2}}}\right)\right] \\ & +\left[\varphi\left(u_{1,qh}\right)\Phi\left(\frac{u_{1,qhK}-\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^{2}}}\right)\right]\Bigg\}X \end{aligned} \tag{4.58}$$

For σ_1, we use transformations on parameters to ensure that in the optimization process, each σ_1 remains non- negative. For σ_1 we use exponential transformation, that is, $d_1 = \log(\sigma_1)$ and $e^{d_1} = \sigma_1$, then we differentiate with respect to d_1 .

$$\frac{\partial \Phi(A,B)}{\partial d_1} = \varphi(A)\frac{\partial \Phi(A,B)}{\partial A}\frac{\partial A}{\partial \sigma_1}\frac{\partial \sigma_1}{\partial d_1}, \ \frac{\partial A}{\partial \sigma_1} = a_h\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)} \tag{4.59}$$

and $\frac{\partial \sigma_1}{\partial d_1} = e^{d_1}$. The differentiation of (4.55) with respect to d_1 is given by,

$$\begin{aligned}\frac{\partial \log f(y_{itk}|\Theta)}{\partial d_1} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\Bigg\{ & \left[-\varphi\left(u_{1,qh}\right)+\varphi\left(u_{1,qh}\right)\Phi\left(\frac{u_{1,qh0}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^{2}}}\right)\right]+ \\ & \left[\varphi\left(u_{1,qh}\right)\Phi\left(\frac{u_{1,qhk}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^{2}}}\right)\right. \\ & \left. -\varphi\left(u_{1,qh}\right)\Phi\left(\frac{u_{1,qhk-1}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^{2}}}\right)\right] \\ & +\left[\varphi\left(u_{1,qh}\right)\Phi\left(\frac{u_{1,qhK}-\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^{2}}}\right)\right]\Bigg\}e^{d_1}a_h\sqrt{2\left(1-\rho^2_{\delta_1\delta_2}\right)} \end{aligned} \tag{4.60}$$

The differentiation of (4.55) with respect to ψ is given by,

$$\frac{\partial \log f\left(y_{itk}|\Theta\right)}{\partial \psi} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[\varphi\left(u_{1,qh0}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qh0}}{\sqrt{1-\rho_{ebeo}^{2}}}\right)\right]+\right.$$
$$\left[\left(\varphi\left(u_{1,qhk}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk}}{\sqrt{1-\rho_{ebeo}^{2}}}\right)\right)-\right.$$
$$\left.\varphi\left(u_{1,qhk-1}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk-1}}{\sqrt{1-\rho_{ebeo}^{2}}}\right)\right]$$
$$\left.+\left[-\varphi\left(u_{1,qhK}\right)\Phi\left(\frac{u_{1,qh}-\rho_{ebeo}u_{1,qhK}}{\sqrt{1-\rho_{ebeo}^{2}}}\right)\right]\right\} \tag{4.61}$$

For σ_2, we use transformations on parameters to ensure that in the optimisation process, each σ_2 remains non-negative. For σ_2 we use exponential transformation, that is, $d_2 = \log(\sigma_2)$ and $e^{d_2} = \sigma_2$ then in the derivation, we differentiate with respect to d_2. The differentiation of (4.55) with respect to d_2 is given by,

$$\frac{\partial \log f\left(y_{itk}|\Theta\right)}{\partial d_2} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[\varphi\left(u_{1,qh0}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qh0}}{\sqrt{1-\rho_{ebeo}^{2}}}\right)\right]+\right.$$
$$\left[\left(\varphi\left(u_{1,qhk}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk}}{\sqrt{1-\rho_{ebeo}^{2}}}\right)\right)-\right.$$
$$\left.\left(\varphi\left(u_{1,qhk-1}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk-1}}{\sqrt{1-\rho_{ebeo}^{2}}}\right)\right)\right]$$
$$\left.+\left[-\varphi\left(u_{1,qhK}\right)\Phi\left(\frac{u_{1,qh}-\rho_{ebeo}u_{1,qhK}}{\sqrt{1-\rho_{ebeo}^{2}}}\right)\right]\right\}$$
$$\left(-a_q e^{d_2}\sqrt{2\left(1-\rho_{\delta_{1i}\delta_{2i}}^{2}\right)}\right) \tag{4.62}$$

Plackett (1954) derived the formula for the partial derivative of the cumulative bivariate normal distribution function with respect to ρ.

$$\frac{\partial \Phi\left(A,\ B,\ \rho\right)}{\partial \rho} = \frac{1}{2\pi\sqrt{1-\rho^{2}}}\exp\left(-\frac{A^{2}+B^{2}-2\rho AB}{2\left(1-\rho^{2}\right)}\right) = \phi\left(A,\ B,\ \rho\right) \tag{4.63}$$

Here, we estimate the additional parameter $r = a\tanh(\rho)$ instead of estimating ρ. We apply Fisher transformation to find ρ on the interval $[-1,\ 1]$ .The differentiation of the bivariate normal distribution function with respect to r is given by,

$$\frac{\partial \Phi(A,B)}{\partial r} = \frac{\partial \Phi(A,B)}{\partial \rho}\frac{\partial \rho}{\partial r} \tag{4.64}$$

where $\frac{\partial \Phi(A,B)}{\partial r} = \frac{\partial \Phi(A,B)}{\partial \rho}\frac{\partial \rho}{\partial r}$ Using (4.63) and (4.64), the differentiation of (3.55) with respect to r_e is given by,

$$\begin{aligned}\frac{\partial \log f(y_{itk}|\Theta)}{\partial r_e} = &\sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[\varphi\left(u_{1,qh},\ u_{1,qh0}, -\rho_{ebeo}\right)\right] + \right.\\ &\left[\left(\varphi\left(u_{1,qh},\ u_{1,qhk}, -\rho_{ebeo}\right)\right) - \left(\varphi\left(u_{1,qh},\ u_{1,qhk-1}, -\rho_{ebeo}\right)\right)\right]\\ &\left.\left[-\varphi\left(u_{1,qh},\ u_{1,qhK},\ \rho_{ebeo}\right)\right]\right\}\frac{-4\exp(2r_e)}{(1+\exp(2r_e))^2}\end{aligned} \tag{4.65}$$

Using 4.64, the differentiation of (4.55) with respect to r is given by,

$$\begin{aligned}
\frac{\partial \log f\left(y_{itk}|\Theta\right)}{\partial r} = & \frac{-\rho_{\delta_{1i}\delta_{2i}}}{\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)} \frac{4\exp(2r)}{\left(1+\exp(2r)\right)^2} + 2a_q a_h \frac{4\exp(2r)}{\left(1+\exp(2r)\right)^2} \\
& + \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b} + \right.\right. \\
& \left(-\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}\Phi\left(\frac{u_{1,qh0}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho^2_{ebeo}}}\right) + \right. \\
& \left.\left.\varphi\left(u_{1,qh0}\right)a_q\sigma_{\delta_o}\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qh0}}{\sqrt{1-\rho^2_{ebeo}}}\right)\right)\right] + \\
& \left[\left(\left(-\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}\Phi\left(\frac{u_{1,qhk}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho^2_{ebeo}}}\right) + \right.\right.\right. \\
& \left.\left.\varphi\left(u_{1,qhk}\right)a_q\sigma_{\delta_o}\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk}}{\sqrt{1-\rho^2_{ebeo}}}\right)\right)\right) - \\
& \left(-\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}\Phi\left(\frac{u_{1,qhk-1}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho^2_{ebeo}}}\right) + \right. \\
& \left.\left.\left.\varphi\left(u_{1,qhk-1}\right)a_q\sigma_{\delta_o}\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}\,u_{1,qhk-1}}{\sqrt{1-\rho^2_{ebeo}}}\right)\right)\right)\right] + \\
& \left[\left(-\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}\Phi\left(\frac{u_{1,qhK}-\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho^2_{ebeo}}}\right) - \right.\right. \\
& \left.\left.\varphi\left(u_{1,qhK}\right)a_q\sigma_{\delta_o}\Phi\left(\frac{u_{1,qh}-\rho_{ebeo}u_{1,qhK}}{\sqrt{1-\rho^2_{ebeo}}}\right)\right)\right] \\
& \frac{\sqrt{2}\rho_{\delta_{1i}\delta_{2i}}}{\sqrt{\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}}\frac{4\exp(2r)}{\left(1+\exp(2r)\right)^2}
\end{aligned} \tag{4.66}$$

To impose the monotone order of the threshold values, we estimate instead a transformation of these values $\alpha_1 = \tau_1,\ \alpha_k = \log\left(\tau_k - \tau_{k-1}\right)$ k = 2, 3, 4, ..., K and $\tau_k = \tau_{k-1} + \mathrm{e}^{\alpha_k}$. The derivatives of the bivariate normal distribution function with respect to α_k is given by,

$$\frac{\partial\Phi(A,B)}{\partial\alpha_k} = \varphi(A)\frac{\partial\Phi(A,B)}{\partial\tau_k}\frac{\partial\tau_k}{\partial\alpha_k} \tag{4.67}$$

where $\frac{\partial\tau_k}{\partial\alpha_k} = \mathrm{e}^{\alpha_k}$ and $\frac{\partial\tau_1}{\partial\alpha_1} = 1$.

Using (4.67), the differentiation of (4.55) with respect to α_2 and α_k are given by,

$$\begin{aligned}\frac{\partial\log f(y_{itk}|\Theta)}{\partial\alpha_1} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\Bigg\{&\left[\varphi\left(u_{1,qh1}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qh1}}{\sqrt{1-\rho^2{}_{ebeo}}}\right)\right]+\\ &\left[-\varphi\left(u_{1,qh1}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qh1}}{\sqrt{1-\rho^2{}_{ebeo}}}\right)\right]\Bigg\}\end{aligned} \tag{4.68}$$

$$\begin{aligned}\frac{\partial\log f(y_{itk}|\Theta)}{\partial\alpha_k} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\Bigg\{&\left[\varphi\left(u_{1,qhk}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk}}{\sqrt{1-\rho^2{}_{ebeo}}}\right)\right]+\\ &\left[-\varphi\left(u_{1,qhk}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk}}{\sqrt{1-\rho^2{}_{ebeo}}}\right)\right]\Bigg\}\mathrm{e}^{\alpha_k}\end{aligned} \tag{4.69}$$

In case the error terms are independent, the first order derivatives are given below. The differentiation of (4.56) with respect to θ was given by,

$$\begin{aligned}\frac{\partial\log f(y_{itk}|\Theta)}{\partial\theta} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\{&\left[-\varphi\left(u_{1,qh}\right)+\varphi\left(u_{1,qh}\right)\Phi\left(u_{1,qh0}\right)\right]+\\ &\left[\varphi\left(u_{1,qh}\right)\left(\Phi\left(u_{1,qhk}\right)-\Phi\left(u_{1,qhk-1}\right)\right)\right]+\left[\varphi\left(u_{1,qh}\right)\Phi\left(u_{1,qhK}\right)\right]\}X\end{aligned} \tag{4.70}$$

The differentiation of (4.56) with respect to d_1 is given by,

$$\begin{aligned}\frac{\partial\log f(y_{itk}|\Theta)}{\partial d_1} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\{&\left[-\varphi\left(u_{1,qh}\right)+\varphi\left(u_{1,qh}\right)\Phi\left(u_{1,qh0}\right)\right]+\\ &\left[\varphi\left(u_{1,qh}\right)\left(\Phi\left(u_{1,qhk}\right)-\Phi\left(u_{1,qhk-1}\right)\right)\right]+\left[\varphi\left(u_{1,qh}\right)\Phi\left(u_{1,qhK}\right)\right]\}\\ &e^{d_1}a_h\sqrt{2\left(1-\rho^2_{\delta_1\delta_2}\right)}\end{aligned} \tag{4.71}$$

The derivative of (4.56) with respect to ψ is given by,

$$\frac{\partial \log f\left(y_{itk}|\Theta\right)}{\partial \psi} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[\varphi\left(u_{1,qh0}\right)\Phi\left(u_{1,qh}\right)\right]+\right.$$
$$\left.\left[\Phi\left(u_{1,qh}\right)\left(\varphi\left(u_{1,qhk}\right)-\varphi\left(u_{1,qhk-1}\right)\right)\right]+\left[-\varphi\left(u_{1,qhK}\right)\Phi\left(u_{1,qh}\right)\right]\right\}Z \tag{4.72}$$

The differentiation of (4.56) with respect to d_2 is given by,

$$\frac{\partial \log f\left(y_{itk}|\Theta\right)}{\partial d_2} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[\varphi\left(u_{1,qh0}\right)\Phi\left(u_{1,qh}\right)\right]+\right.$$
$$\left.\left[\Phi\left(u_{1,qh}\right)\left(\varphi\left(u_{1,qhk}\right)-\varphi\left(u_{1,qhk-1}\right)\right)\right]+\left[-\varphi\left(u_{1,qhK}\right)\Phi\left(u_{1,qh}\right)\right]\right\}$$
$$\left(-a_q e^{d_2}\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\right) \tag{4.73}$$

Using (4.64), the differentiation of (4.56) with respect to r is given by,

$$\frac{\partial \log f\left(y_{itk}|\Theta\right)}{\partial r} = \frac{-\rho_{\delta_{1i}\delta_{2i}}}{\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\frac{4\exp(2r)}{\left(1+\exp(2r)\right)^2} + 2a_q a_h \frac{4\exp(2r)}{\left(1+\exp(2r)\right)^2}$$
$$+\sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}+\right.\right.$$
$$\left.\left(-\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}\Phi\left(u_{1,qh0}\right)+\varphi\left(u_{1,qh0}\right)a_q\sigma_{\delta_o}\Phi\left(u_{1,qh}\right)\right)\right]$$
$$\left[-\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}\left(\Phi\left(u_{1,qhk}\right)+\Phi\left(u_{1,qhk-1}\right)\right)\right.$$
$$\left.+\Phi\left(u_{1,qh}\right)\left(-\varphi\left(u_{1,qhk}\right)a_q\sigma_{\delta_o}+\varphi\left(u_{1,qhk-1}\right)a_q\sigma_{\delta_o}\right)\right]$$
$$\left.+\left[-\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}\Phi\left(u_{1,qhK}\right)-\varphi\left(u_{1,qhK}\right)a_q\sigma_{\delta_o}\Phi\left(u_{1,qh}\right)\right]\right\}$$
$$\frac{\sqrt{2}\rho_{\delta_{1i}\delta_{2i}}}{\sqrt{\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}}\frac{4\exp(2r)}{\left(1+\exp(2r)\right)^2} \tag{4.74}$$

Using equation (4.67), the differentiation of (4.56) with respect to α_2 and α_k are given by,

$$\frac{\partial \log f\left(y_{itk}|\Theta\right)}{\partial \alpha_1} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[\varphi\left(u_{1,qh1}\right)\Phi\left(u_{1,qh}\right)\right]+\left[-\varphi\left(u_{1,qh1}\right)\Phi\left(u_{1,qh}\right)\right]\right\} \tag{4.75}$$

$$\frac{\partial \log f\left(y_{itk}|\Theta\right)}{\partial \alpha_k} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[\varphi\left(u_{1,qhk}\right)\Phi\left(u_{1,qh}\right)\right] + \left[-\varphi\left(u_{1,qhk}\right)\Phi\left(u_{1,qh}\right)\right]\right\}\mathrm{e}^{\alpha_k} \tag{4.76}$$

The second order derivatives are given in the appendix.

4.5 Model selection

The study evaluated the statistical fit of the models using AIC for model selection. Formally,

$$AIC = -2logL + 2r \tag{4.77}$$

where r denoted the number of estimates and $logL$ the maximized log-likelihood function. The model with the smallest AIC is preferred.

CHAPTER FIVE

BAYESIAN APPROACH FOR ZERO INFLATION DYNAMIC PANEL ORDERED PROBIT MODEL

5.1 Introduction

This chapter constitutes a discussion on Bayesian analysis, model selection and Average partial effects ZIDPOP models.

5.2 Bayesian Analysis

The Bayesian estimation based on MCMC simulation depends on a collection of conditional distributions to infer each parameter's marginal distribution. Therefore, models with a lot of parameters and intricated multiple-layered probability specifications can be disintegrated into a collection of easier sub-problems. In addition, Bayesian approach permit priors knowledge, recognized intuition and experience in the inference.

The posterior distribution will for most models not have a closed-form analytical expression, but rather be a high dimensional integral. The development over the last decades of highly efficient Markov chain Monte Carlo simulation algorithms combined with increased access to inexpensive high-speed computing have made Bayesian statistical estimation possible for a large variety of distributions, and furthermore allowed estimation of models not tractable or feasible with classical statistical tools.

The ZIDPOP models described earlier naturally leads to Bayesian modelling. Sampling from the posterior allows us to make predictions while taking into account parameter uncertainty. Bayesian analysis can easily incorporate truncated distributions and allows us to use subjective information in our priors.

Conditioned on the unobserved heterogeneity κ, the responses on $y_{it}\,(t = 1,\ 2,\ ...,\ T)$, are considered uncorrelated. Using equation (4.24) and (4.25), the contribution to the likelihood for respondent i, conditional on the explanatory variables and the unobserved heterogeneities, would be the joint probability

$$
\begin{aligned}
p\left(y_{it}|others\right) = & \prod_{i=1}^{n}\prod_{t=1}^{T}\prod_{k=0} \\
& \left[\left(1-\Phi\left(\phi_1 y_{it-1}^{b}+\gamma'_1 x_{it}+\beta'_1 w_i+h_0^b+h_1^b y_{i0}^{b}+h_2^b\overline{x}_i+\delta_{1i}\right)\right)+\right. \\
& \Phi\left(\phi_1 y_{it-1}^{b}+\gamma'_1 x_{it}+\beta'_1 w_i+h_0^b+h_1^b y_{i0}^{b}+h_2^b\overline{x}_i+\delta_{1i},\right. \\
& \left.\left.-\phi_2 y_{it-1}^{o}-\gamma'_2 z_{it}-\beta'_2 v_i-h_0^o-h_1^o y_{i0}^{0}-h_2^o\overline{z}_i-\delta_{2i},\ -\rho\right)\right] \\
& \prod_{i=1}^{n}\prod_{t=1}^{T}\prod_{k>0}^{K}\left[\Phi\left(\phi_1 y_{it-1}^{b}+\gamma'_1 x_{it}+\beta'_1 w_i+h_0^b+h_1^b y_{i0}^{b}+h_2^b\overline{x}_i+\delta_{1i}\right.\right. \\
& \left.,\ \tau_k-\phi_2 y_{it-1}^{o}-\gamma'_2 z_{it}-\beta'_2 v_i-h_0^o-h_1^o y_{i0}^{0}-h_2^o\overline{z}_i-\delta_{2i},\ -\rho\right) \\
& -\Phi\left(\phi_1 y_{it-1}^{b}+\gamma'_1 x_{it}+\beta'_1 w_i+h_0^b+h_1^b y_{i0}^{b}+h_2^b\overline{x}_i+\delta_{1i},\right. \\
& \left.\left.\tau_{k-1}-\phi_2 y_{it-1}^{o}-\gamma'_2 z_{it}-\beta'_2 v_i-h_0^o-h_1^o y_{i0}^{0}-h_2^o\overline{z}_i-\delta_{2i},\ -\rho\right)\right]
\end{aligned}
\tag{5.1}
$$

where $\Phi(.)$ is the cumulative standard normal distribution.

$$
\begin{pmatrix} e_{it}^{b} \\ e_{it}^{o} \end{pmatrix} \sim N\left(0,\ \Sigma_{ebeo}\right)
$$

where

$$
\Sigma_{ebeo} = \begin{pmatrix} 1 & \rho_{ebeo} \\ \rho_{ebeo} & 1 \end{pmatrix}
$$

$$
\begin{pmatrix} \kappa_{1i}^{b} \\ \kappa_{2i}^{o} \end{pmatrix} \sim N\left(0,\ \Sigma_{\delta_1\delta_2}\right)
$$

where

$$
\Sigma_{\delta_1\delta_2} = \begin{pmatrix} \sigma_{\delta_1}^2 & \rho_{\delta_1\delta_2}\sigma_{\delta_1}\sigma_{\delta_2} \\ \rho_{\delta_1\delta_2}\sigma_{\delta_1}\sigma_{\delta_2} & \sigma_{\delta_2}^2 \end{pmatrix}
$$

be correlation and covariance matrix for errors terms at initial and waves respectively and unobserved effect between the binary and ordinal variables.

5.3 Prior Distributions

Let Θ denote a vector of parameters to be estimated and is given by $\Theta = \left\{\phi_1,\ \phi_2,\ \gamma_1,\ \beta_1,\ \gamma_2,\ \beta_2,\ \Sigma_{ebeo},\ \Sigma_{\delta_1\delta_2},\ \tau\right\}$. The prior distribution of κ was considered to possess a hierarchical structure. Then, assumming independence, the hierarchical structure is given by

$$\begin{aligned} p(\Theta) = {} & p(\phi_1)\, p(\phi_2)\, p(\gamma_1)\, p(\beta_1)\, p(\gamma_2)\, p(\beta_2) * \\ & p\left(\kappa_1 | 0,\ \Sigma_{\delta_1\delta_2}\right) p\left(\Sigma_{\delta_1\delta_2}\right) p\left(\kappa_2 | 0,\ \Sigma_{\delta_1\delta_2}\right) p(\tau)\, p\left(\Sigma_{ebeo}\right) \end{aligned} \tag{5.2}$$

The study adopted a Bayesian hierarchical model approach for estimation. The prior specifications for most of the parameters are non-informative. The motive for taking non-informative priors is due to least influence on inference.

5.4 Posterior Distributions

The posterior distributions are constructed by coalescing the prior distributions and likelihood function based on Bayes' theorem, as:

$$\begin{aligned} p(\Theta | y_{it}, x, z, w, v) = {} & p(\Theta) \prod_{i=1}^{n} \prod_{t=1}^{T} \prod_{k=0} \\ & \Big[\Big(1 - \Phi\Big(\phi_1 y^b_{it-1} + \gamma'_1 x_{it} + \beta'_1 w_i + h^b_0 + h^b_1 y^b_{i0} + h^b_2 \bar{x}_i + \delta_{1i}\Big)\Big) + \\ & \Phi\Big(\phi_1 y^b_{it-1} + \gamma'_1 x_{it} + \beta'_1 w_i + h^b_0 + h^b_1 y^b_{i0} + h^b_2 \bar{x}_i + \delta_{1i}, \\ & -\phi_2 y^o_{it-1} - \gamma'_2 z_{it} - \beta'_2 v_i - h^o_0 - h^o_1 y^0_{i0} - h^o_2 \bar{z}_i - \delta_{2i},\ -\rho\Big)\Big] \\ & \prod_{i=1}^{n} \prod_{t=1}^{T} \prod_{k>0}^{K} \Big[\Phi\Big(\phi_1 y^b_{it-1} + \gamma'_1 x_{it} + \beta'_1 w_i + h^b_0 + h^b_1 y^b_{i0} + h^b_2 \bar{x}_i + \delta_{1i}, \\ & \tau_k - \phi_2 y^o_{it-1} - \gamma'_2 z_{it} - \beta'_2 v_i - h^o_0 - h^o_1 y^0_{i0} - h^o_2 \bar{z}_i - \delta_{2i},\ -\rho\Big) \\ & -\Phi\Big(\phi_1 y^b_{it-1} + \gamma'_1 x_{it} + \beta'_1 w_i + h^b_0 + h^b_1 y^b_{i0} + h^b_2 \bar{x}_i + \delta_{1i}, \\ & \tau_{k-1} - \phi_2 y^o_{it-1} - \gamma'_2 z_{it} - \beta'_2 v_i - h^o_0 - h^o_1 y^0_{i0} - h^o_2 \bar{z}_i - \delta_{2i},\ -\rho\Big)\Big] \end{aligned} \tag{5.3}$$

We assumed non-informative normal priors for $\theta = (\phi_1,\ \phi_2,\ \gamma_1,\ \gamma_2,\ \beta_1,\ \beta_2)$, with mean θ^* and variance Ω_θ , that is, $\theta \sim N(\theta^*,\ \Omega_\theta)$ that are picked to produce proper distribution, that is, integrates to 1 but diffuse with large variances. ,

$$p(\theta_t) \propto \Omega_\theta^{-1/2} \exp\left\{-1/2(\theta-\theta^*)'\Omega_\theta^{-1}(\theta-\theta^*)\right\} \tag{5.4}$$

5.5 Full conditional distributions

The Metropolis-Hastings algorithm was used due to the difficult to sample from the unknown full conditional distributions. The full conditional distributions are given by the product of likelihood function and prior distribution. The full conditional posterior distributions to implement the MCMC algorithm are given by,

$$\begin{aligned} p(\theta|x,z,w,v,\theta_\theta) \propto {} & |\Omega_\theta|^{-\frac{1}{2}} exp\left(-\frac{1}{2}(\theta-\theta^*)'\Omega_\theta^{-1}(\theta-\theta^*)\right)\prod_{i=1}^{n}\prod_{t=1}^{T}\prod_{k=0} \\ & \left[\left(1-\Phi\left(\phi_1 y_{it-1}^b+\gamma'_1 x_{it}+\beta'_1 w_i+h_0^b+h_1^b y_{i0}^b+h_2^b\bar{x}_i+\delta_{1i}\right)\right)\right. \\ & +\Phi\left(\phi_1 y_{it-1}^b+\gamma'_1 x_{it}+\beta'_1 w_i+h_0^b+h_1^b y_{i0}^b+h_2^b\bar{x}_i+\delta_{1i},\right. \\ & \left.\left.-\phi_2 y_{it-1}^o-\gamma'_2 z_{it}-\beta'_2 v_i-h_0^o-h_1^o y_{i0}^0-h_2^o\bar{z}_i-\delta_{2i},-\rho\right)\right] \\ & \prod_{i=1}^{n}\prod_{t=1}^{T}\prod_{k>0}^{K}\left[\Phi\left(\phi_1 y_{it-1}^b+\gamma'_1 x_{it}+\beta'_1 w_i+h_0^b+h_1^b y_{i0}^b+h_2^b\bar{x}_i+\delta_{1i},\right.\right. \\ & \left.\tau_k-\phi_2 y_{it-1}^o-\gamma'_2 z_{it}-\beta'_2 v_i-h_0^o-h_1^o y_{i0}^0-h_2^o\bar{z}_i-\delta_{2i},\ -\rho\right) \\ & -\Phi\left(\phi_1 y_{it-1}^b+\gamma'_1 x_{it}+\beta'_1 w_i+h_0^b+h_1^b y_{i0}^b+h_2^b\bar{x}_i+\delta_{1i},\right. \\ & \left.\left.\tau_{k-1}-\phi_2 y_{it-1}^o-\gamma'_2 z_{it}-\beta'_2 v_i-h_0^o-h_1^o y_{i0}^0-h_2^o\bar{z}_i-\delta_{2i},\ -\rho\right)\right] \end{aligned} \tag{5.5}$$

The normal distributions applied as the proposal distributions are symmetric. The probability of acceptance is given by,

$$\alpha\left(x^{(i)}|x^{(i-1)}\right) = \min\left(1,\ \frac{\pi\left(x^{(i)}\right)}{\pi\left(x^{(i-1)}\right)}\right) \tag{5.6}$$

Since $q\left(x^{(i)}|x^{(i-1)}\right) = q\left(x^{(i-1)}|x^{(i)}\right)$ for symmetrical distribution. Let $u \sim U(0,\ 1)$. If $u < \alpha$ then admit the proposal: $x^{(i)} \leftarrow x^{candidate}$ (admit a newly proposed value) else fail to admit the proposal: $x^{(i)} \leftarrow x^{(t-1)}$.

The prior for variance –covariance matrix

$$\Sigma_{\delta_1\delta_2} = \begin{pmatrix} \sigma^2_{\delta_1} & \rho_{\delta_1\delta_2}\sigma_{\delta_1}\sigma_{\delta_2} \\ \rho_{\delta_1\delta_2}\sigma_{\delta_1}\sigma_{\delta_2} & \sigma^2_{\delta_2} \end{pmatrix}$$

was inverse Wishart distributions denoted by $\Sigma \sim IW(v, \Lambda)$ with density.

$$p(\Sigma) = \frac{|\Lambda|^{\frac{v}{2}}}{|\Sigma|^{\frac{v+d+1}{2}} 2^{\frac{vd}{2}} \Gamma_d\left(\frac{v}{2}\right)} \exp\left(-\frac{1}{2} tr\left(\Lambda\Sigma^{-1}\right)\right) \tag{5.7}$$

$$p(\Sigma) \propto |\Sigma|^{-\frac{v+d+1}{2}} \exp\left(-\frac{1}{2} tr\left(\Lambda\Sigma^{-1}\right)\right) \tag{5.8}$$

Λ is 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 variance-covariance matrix induces an inverse scale chi-square distribution for each variance $\sigma_i \sim inv\chi^2\left(v-d+1,\ \frac{\lambda_{ii}}{v-d+1}\right)$ where λ_{ii} it is a diagonal entry of Λ. $\lambda_{ii}$is obtained by decomposing Σ into a diagonal matrix with standard deviation and correlation matrix with diagonal elements 1 and off-diagonal elements ρ_{ij}. The conditional distribution of the correlation coefficient is given by

$$p\left(\rho|\sigma_1^2, \sigma_2^2\right) \propto \left(1-\rho^2\right)^{-\frac{v+d+1}{2}} \exp\left(-\frac{\lambda_{22}\sigma_1^2 + \lambda_{11}\sigma_2^2 - 2\lambda_{12}\sigma_1\sigma_2}{\sigma_1^2\sigma_2^2\left(1-\rho^2\right)}\right) \tag{5.9}$$

Assuming an $IW(v_0,\ D_0)$ where $v_0 = 2$ and $D_0 = I$, I is an identity matrix, draw $\Sigma_{\delta_1\delta_2}$ from its full conditional

$$\begin{aligned} \pi\left(\Sigma_{\delta_1\delta_2}|others\right) &\propto \pi\left(\Sigma_{\delta_1\delta_2}|\kappa_i\right) \\ &\propto IW\left(n+v_0,\ \mathrm{D}_0 + \kappa'_i\kappa_i\right) \end{aligned} \tag{5.10}$$

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

$$\begin{pmatrix} \kappa_{1i}^{b} \\ \kappa_{2i}^{o} \end{pmatrix} \sim N\left(0,\ \Sigma_{\delta_1\delta_2}\right)$$

where

$$\Sigma_{\delta_1\delta_2} = \begin{pmatrix} \sigma_{\delta_1}^2 & \rho_{\delta_{1i}\delta_{2i}}\sigma_{\delta_1}\sigma_{\delta_2} \\ \rho_{\delta_{1i}\delta_{2i}}\sigma_{\delta_1}\sigma_{\delta_2} & \sigma_{\delta_2}^2 \end{pmatrix}$$

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 $h_\kappa(\kappa)$, from which proposed values are sampled. If we choose f_κ that is in the exponential family as the proposal distribution then the acceptance function assumes a particularly an orderly form. Let κ denote the preceding draw from the conditional distribution of $\kappa|y$ and propose a value, κ^* , for the k^{th} component of κ using the proposal distribution. If we represent $\kappa^* = (\kappa_1,\ \kappa_2,\ ...,\ \kappa_{i-1},\ \kappa_i^*,\ \kappa_{i+1},\ ...,\ \kappa_i)$, next we admit κ^* as the new value with probability $\alpha_i(\kappa,\ \kappa^*)$ and otherwise we retain κ. Here $\alpha_i(\kappa,\ \kappa^*)$ is given by

$$\alpha_i(\kappa,\ \kappa^*) = \min\left(1, \frac{f_{\kappa|y}(\kappa^*|y,\ others)\,h_\kappa(\kappa)}{f_{\kappa|y}(\kappa|y,\ others)\,h_\kappa(\kappa^*)}\right) \tag{5.11}$$

Upon choosing $h_\kappa = f_k$, the second term in braces in the above equation simplifies to

$$\begin{aligned} \frac{f_{\kappa|y}(\kappa^*|y,\ others)\,h_\kappa(\kappa)}{f_{\kappa|y}(\kappa|y,\ others)\,h_\kappa(\kappa^*)} &= \frac{\prod_{i=1}^{n}\prod_{t=1}^{T} f_{y|\kappa}(y_{it}|\kappa^*, others)\, f_\kappa\left(\kappa^*|\Sigma_{\delta_1\delta_2}\right) f_\kappa\left(\kappa|\Sigma_{\delta_1\delta_2}\right)}{\prod_{i=1}^{n}\prod_{t=1}^{T} f_{y|\kappa}(y_{it}|\kappa, others)\, f_\kappa\left(\kappa|\Sigma_{\delta_1\delta_2}\right) f_\kappa\left(\kappa^*|\Sigma_{\delta_1\delta_2}\right)} \\ &= \frac{\prod_{i=1}^{n}\prod_{t=1}^{T} f_{y|\kappa}(y_{it}|\kappa^*, others)}{\prod_{i=1}^{n}\prod_{t=1}^{T} f_{y|\kappa}(y_{it}|\kappa, others)} \end{aligned} \tag{5.12}$$

The only unknown parameter in Θ is ρ, the correlation between e_{it}^{b} and e_{it}^{o}. The values of ρ lies within -1 to 1 interval. Uniform distribution or a proper distribution based on reparametrization are the choices for prior distribution of ρ. Assume v represent the hyperbolic arc-tangent transformation of the correlation coefficient, that is,

$$v = a\tanh(\rho) \tag{5.13}$$

and using hyperbolic tangent transformation of v return $\rho = \tanh(v)$. We consider $v \sim N\left(v^{*}, \sigma_v^2\right)$. The full conditional posterior distribution to implement the MCMC algorithm is given by,

$$\begin{aligned}
f(v|others) \propto \sigma_v^{-1} \exp\left\{-\frac{(v-v^{*})}{2\sigma_v^2}\right\} \prod_{i=1}^{n}\prod_{t=1}^{T}\prod_{k=0} \\
\Big[\Big(1-\Phi\Big(\phi_1 y_{it-1}^{b} + \gamma'_1 x_{it} + \beta'_1 w_i + h_0^b + h_1^b y_{i0}^b + h_2^b \bar{x}_i + \delta_{1i}\Big)\Big) \\
+\Phi\Big(\phi_1 y_{it-1}^{b} + \gamma'_1 x_{it} + \beta'_1 w_i + h_0^b + h_1^b y_{i0}^b + h_2^b \bar{x}_i + \delta_{1i}, \\
-\phi_2 y_{it-1}^{o} - \gamma'_2 z_{it} - \beta'_2 v_i - h_0^o - h_1^o y_{i0}^0 - h_2^o \bar{z}_i - \delta_{2i}, -\rho\Big)\Big] \\
\prod_{i=1}^{n}\prod_{t=1}^{T}\prod_{k>0}^{K}\Big[\Phi\Big(\phi_1 y_{it-1}^{b} + \gamma'_1 x_{it} + \beta'_1 w_i + h_0^b + h_1^b y_{i0}^b + h_2^b \bar{x}_i + \delta_{1i}, \\
\tau_k - \phi_2 y_{it-1}^{o} - \gamma'_2 z_{it} - \beta'_2 v_i - h_0^o - h_1^o y_{i0}^0 - h_2^o \bar{z}_i - \delta_{2i},\ -\rho\Big) \\
-\Phi\Big(\phi_1 y_{it-1}^{b} + \gamma'_1 x_{it} + \beta'_1 w_i + h_0^b + h_1^b y_{i0}^b + h_2^b \bar{x}_i + \delta_{1i}, \\
\tau_{k-1} - \phi_2 y_{it-1}^{o} - \gamma'_2 z_{it} - \beta'_2 v_i - h_0^o - h_1^o y_{i0}^0 - h_2^o \bar{z}_i - \delta_{2i},\ -\rho\Big)\Big]
\end{aligned} \tag{5.14}$$

The proposal distribution for each of the variables of interest given above is normal distribution that is a symmetric distribution. When selecting the prior distributions for the cut points, $\tau' s$, care is required because of the order constraints. Drawing values from the complete conditional distribution for $\tau' s$ leads to slow mixing. Instead, we apply Cowles (1996) algorithm to simulate $\tau' s$. Let $i = 1$ and $\sigma_{MH} = 0.05/K$. This value of σ_{MH} is a starting, and tuning of σ_{MH} may be required if suitable acceptance rates for τ are not attained. Create a candidate g for bringing up-to-date τ^{k-1} : For $j = 1, \ldots, K-1$, sample $g_j \sim N\left(\tau_j^{i-1}, \sigma_{MH}^2\right)$ trimmed to the interval $\left(g_{j-1},\ \tau_{j+1}^{i-1}\right)$

$\left(take\ g_{-1}=-\infty,\ g_0=0\ and\ g_K=\infty\right)$. Compute the acceptance rate R by

$$R=\prod_{i=1}^{n}\prod_{t=1}^{T}\Big[\Phi\Big(\phi_1 y^b_{it-1}+\gamma'_1 x_{it}+\beta'_1 w_i+h^b_0+h^b_1 y^b_{i0}+h^b_2\bar{x}_i+\delta_{1i},$$
$$g_{y_i}-\phi_2 y^o_{it-1}-\gamma'_2 z_{it}-\beta'_2 v_i-h^o_0-h^o_1 y^0_{i0}-h^o_2\bar{z}_i-\delta_{2i},\ -\rho_{ebeo}\Big)-$$
$$\Phi\Big(\phi_1 y^b_{it-1}+\gamma'_1 x_{it}+\beta'_1 w_i+h^b_0+h^b_1 y^b_{i0}+h^b_2\bar{x}_i+\delta_{1i},$$
$$g_{y_i-1}-\phi_2 y^o_{it-1}-\gamma'_2 z_{it}-\beta'_2 v_i-h^o_0-h^o_1 y^0_{i0}-h^o_2\bar{z}_i-\delta_{2i},\ -\rho_{ebeo}\Big)\Big]\Big/$$
$$\Big[\Phi\Big(\phi_1 y^b_{it-1}+\gamma'_1 x_{it}+\beta'_1 w_i+h^b_0+h^b_1 y^b_{i0}+h^b_2\bar{x}_i+\delta_{1i},$$
$$\tau_{y_i}-\phi_2 y^o_{it-1}-\gamma'_2 z_{it}-\beta'_2 v_i-h^o_0-h^o_1 y^0_{i0}-h^o_2\bar{z}_i-\delta_{1i},\ -\rho_{ebeo}\Big)-$$
$$\Phi\Big(\phi_1 y^b_{it-1}+\gamma'_1 x_{it}+\beta'_1 w_i+h^b_0+h^b_1 y^b_{i0}+h^b_2\bar{x}_i+\delta_{1i},$$
$$\tau_{y_i-1}-\phi_2 y^o_{it-1}-\gamma'_2 z_{it}-\beta'_2 v_i-h^o_0-h^o_1 y^0_{i0}-h^o_2\bar{z}_i-\delta_{2i},\ -\rho_{ebeo}\Big)\Big]$$
$$\prod_{j=1}^{K-1}\frac{\Phi\left(\left(\tau^{i-1}_{j+1}-\tau^{i-1}_{j}\right)/\sigma_{MH}\right)-\Phi\left(\left(g_{j-1}-\tau^{i-1}_{j}\right)/\sigma_{MH}\right)}{\Phi\left(\left(g_{j+1}-g_j\right)/\sigma_{MH}\right)-\Phi\left(\left(\tau^{i-1}_{j-1}-g_j\right)/\sigma_{MHMH}\right)} \tag{5.15}$$

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 denotes difference in the normalization of the proposal densities on the trimmed normal intervals from which candidate points are sampled. When implementing this algorithm, the acceptance rate for the cut points must be monitored. When the rate are below 25% or above 50%, σ_{MH} should be raised or lowered, respectively.

A single long chain as proposed by Geyer (1992) was applied for the developed models. Geyer (1992) claimed that applying a single longer chain is superior than applying a number of smaller chains with different initial values. We applied this scheme in our study.

5.6 Model Selection

Complex models normally offer a better fit. AIC selection procedures include "penalty" part to compensate for gains in model fit due to increased complexity. For fixed effect models, the complexity–as evaluated by the number of model parameters–is determined without difficulty. It is easier to determine the complexity as measured by the number of model parameters in the fixed effect model as it is easier to count the parameters. This is infeasible in a random effect models as the number of parameters is not well-defined. Every random effect will add roughly one parameter to the model when it have large variance. However, the number of parameter added will be near zero when it have small variance (implying a huge amount of shrinkage). Spiegelhalter *et al.* (2002) dealt with this hurdle by proposing a Deviance Information Criterion given by, goodness of fit plus a penalty for complexity

$$DIC = \overline{D}(\theta) + p_D \tag{5.16}$$

where $\overline{D}(\theta) = E(D(\theta)|y)$ is the posterior mean of the deviance, $D(\theta)$, and $p_D = \overline{D}(\theta) - \hat{D}(\theta) = E[D(\theta)|y] - D[E(\theta|y)]$ is the difference in the posterior mean of the deviance and the deviance computed at the posterior mean of the parameters.

The deviance is given by the negative twice the log-likelihood is a measure of the model's relative fit, whereas p_D represents a penalty for the model's complexity. The model with the smallest DIC provide the best fit. The DIC is easily obtained from the MCMC samples. DIC is estimated from posterior samples:

$$DIC = 2\overline{D} - D\left(\overline{\theta}\right) \tag{5.17}$$

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

5.7 Average Partial Effects

Interpretations of the coefficients in the zero inflated dynamic panel ordered probit models are more difficult than in the usual regression setting. The natural conditional mean function in the model is unavailable.

The response, y_{it}, is just a brand for the binary or ordinal outcomes. The size of the coefficients are not informative and therefore the exact inference of the coefficients is basically unclear. The influence of a shift in a single variable relies on all the model parameters, the data, and which probability (cell) is of interest. The interpretation of the parameters relies on the probabilities themselves. The hurdle is solved by the average partial effects that deliver the impacts on the specific probabilities per unit change in the covariates. Average partial effects gives a good estimation of the amount of change in dependent variable that will be yielded by a single-unit change in covariates. In case of binary independent variables, the average partial effects measure discrete change, i.e. how do predicted probabilities shift as the binary covariate shift from 0 to 1?

Average partial effects are obtained by averaging the individual marginal effects and gives a clue on the magnitude of the relationship between ordinal observation and the covariates. They are evaluated by scaling the coefficient vector by $\gamma_a = \frac{1}{\sqrt{1+\sigma^2}}$. This rescaling is needed in a random effect probit model in order to make valid comparisons in terms of coefficient estimates and partial effects across different specification as pointed out by Arulampalam (1999). A positive average partial effects implying a positive relationship with ordered response and vice versa. The study has obtain the average partial effects for the covariate $\omega = (x_i,\, x_{it},\, z_i,\, z_{it})$ on a range of probabilities assuming the errors terms have a bivariate normal distribution. A positive coefficient means that an increase in the predictor lead to an increase in the predicted probability and vice versa.

For a binary covariate such as w_i, its average partial effects on probability, say P, is the difference in the probability computed at 1 and 0, conditional on observable values of covariates, that is,

$$
\begin{aligned}
APE_i\left(\beta, y_{it}^{b*} = 1\right) = \frac{\beta}{NT} \times \\
\sum_{I=1}^{N}\sum_{t=1}^{T}\Big\{\Phi\left(\phi_1 y_{it-1}^{b} + \gamma'_1 x_{it} + \beta'_1 1 + h_0^b + h_1^b y_{i0}^{b} + h_2^b \bar{x}_i + h_3^b x_{i0}\right) \\
-\Phi\left(\phi_1 y_{it-1}^{b} + \gamma'_1 x_{it} + \beta'_1 0 + h_0^b + h_1^b y_{i0}^{b} + h_2^b \bar{x}_i + h_3^b x_{i0} + h_3^b x_{i0}\right)
\end{aligned}
\tag{5.18}
$$

For continuous covariates, the average partial effects is given by the partial derivative of the probability of interest with respect to ω_i. Wooldridge (2002) showed that evaluating the average partial effects at the observed values of the covariates for each observation and averaging the estimates over the observations provides a consistent estimate of the

APEs.

$$APE_i\left(\lambda_{\omega_i}, y_{it}^{b*}=1\right)=\frac{\lambda_{\omega_i}}{NT}\times \sum_{I=1}^{N}\sum_{t=1}^{T}\varphi\left(\phi_1 y_{it-1}^{b}+\gamma'_1 x_{it}+\beta'_1 w_i+h_0^b+h_1^b y_{i0}^b+h_2^b\bar{x}_i+h_3^b x_{i0}\right) \tag{5.19}$$

where λ_{ω_i} is the coefficient in the inflation part related to variable ω_i. In terms of the zeros response, the effect on the probability of non-participation and zero consumption were,

$$APE_i\left(\lambda_{\omega_i}, y_{it}^{b*}=0\right)=\frac{\lambda_{\omega_i}}{NT}\sum_{I=1}^{N}\sum_{t=1}^{T}-\varphi\left(u_{1,qh}\right) \tag{5.20}$$

while

$$APE_i\left(\lambda_{\omega_i},\ \beta_{\omega_i},\ y_{it}^{b*}=1, y_{it}^{o*}=0\right)= \frac{1}{NT}\sum_{I=1}^{N}\sum_{t=1}^{T}\left(\left[\left[\Phi\left(\frac{-u_{1,qh0}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right]\varphi\left(u_{1,qh}\right)\lambda_{\omega_i} - \left[\Phi\left(\frac{u_{1,qh}-\rho_{ebeo}u_{1,qh0}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right]\varphi\left(-u_{1,qh0}\right)\beta_{\omega_i}\right) \tag{5.21}$$

The average partial effects for zero inflation was the sum of equation (5.20) and (5.21); that is,

$$APE_i\left(y_{it}^{o*}=0\right)=APE_i\left(y_{it}^{b*}=0\right)+APE_i\left(y_{it}^{b*}=1, y_{it}^{o*}=0\right) \tag{5.22}$$

The effects for the remaining choices are: Let $u_{1,qhk}=\tau_k-W,\ u_{1,qhk-1}=\tau_{k-1}-W$ and $u_{1,qhK}=W-\tau_{K-1}$. $W=\phi_2 y_{it-1}^{o}+\gamma'_2 z_{it}+\beta_2 v_i+h_0^o+h_1^o y_{i0}^0+h_2^o z_i+h_3^o z_{i0}$

$$APE_i\left(\lambda_{\omega_i},\ \beta_{\omega_i},\ y_{it}^{b*}=1, y_{it}^{o*}=1\right)=$$
$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\Phi\left(\frac{(\tau_2-W)+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)-\Phi\left(\frac{-W+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right]\varphi\left(u_{1,qh}\right)\lambda_{\omega_i}$$
$$-\left[\Phi\left(\frac{u_{1,qh}++\rho_{ebeo}\left(\tau_2-W\right)}{\sqrt{1-\rho_{ebeo}^2}}\right)\varphi\left(\tau_2-W\right)-\Phi\left(\frac{u_{1,qh}-\rho_{ebeo}W}{\sqrt{1-\rho_{ebeo}^2}}\right)\varphi\left(-W\right)\right]\beta_{\omega_i} \tag{5.23}$$

$$APE_i\left(\lambda_{\omega_i},\ \beta_{\omega_i},\ y_{it}^{b*}=1, y_{it}^{o*}=2\right)=$$
$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\Phi\left(\frac{(\tau_3-W)+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)-\Phi\left(\frac{\tau_2-W+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right]\varphi\left(u_{1,qh}\right)\lambda_{\omega_i}-$$
$$\left[\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}\left(\tau_3-W\right)}{\sqrt{1-\rho_{ebeo}^2}}\right)\varphi\left(\tau_3-W\right)-\right.$$
$$\left.\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}\left(\tau_2-W\right)}{\sqrt{1-\rho_{ebeo}^2}}\right)\varphi\left(\tau_2-W\right)\right]\beta_{\omega_i} \tag{5.24}$$

$$APE_i\left(\lambda_{\omega_i},\ \beta_{\omega_i},\ y_{it}^{b*}=1, y_{it}^{o*}=k\right)=$$
$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\Phi\left(\frac{(\tau_k-W)+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)-\Phi\left(\frac{\tau_{k-1}-W+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right]\varphi\left(u_{1,qh}\right)\lambda_{\omega_i}-$$
$$\left[\Phi\left(\frac{u_{1,qh}++\rho_{ebeo}\left(\tau_k-W\right)}{\sqrt{1-\rho_{ebeo}^2}}\right)\varphi\left(\tau_j-W\right)-\right.$$
$$\left.\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}\left(\tau_{k-1}-W\right)}{\sqrt{1-\rho_{ebeo}^2}}\right)\varphi\left(\tau_{k-1}-W\right)\right]\beta_{\omega_i} \tag{5.25}$$

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

$$\varphi\left(u_{1,qh}\right)\lambda_{\omega_i}+\left[\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}\left(\tau_K-W\right)}{\sqrt{1-\rho_{ebeo}^2}}\right)\varphi\left(\tau_K-W\right)\right]\beta_{\omega_i} \quad (5.26)$$

The asymptotic standard errors of the APE_i are evaluated using the Delta method as the square roots of the main diagonal elements of

$$Var\left(\widehat{APE}\right)=APE_i Var\left(\Theta\right) APE_i' \quad (5.27)$$

$Var\left(\Theta\right)$ represents the variance-covariance matrix.

CHAPTER SIX

RESULTS AND DISCUSSIONS

6.1 Introduction

This chapter introduces the data simulation, analysis, summary of the real data and the finding for both simulation and real data.

6.2 Data Simulation

This section describe the data generating process that was used to simulate the data. In order to generate the first observation, the study operated the scheme twenty-five periods before data were observed. The number of the respondents was (n = 750), (s = -25) represent the time the system was in operation before the first observation was made and the number of observations for the present simulation, so that each respondent was observed from -25 to 36 periods such as months. The study generated the covariates, the unobserved heterogeneities 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 data creating processes for the latent variables were centered on a dynamic random-effects specification given by equation 5.1 and 5.7. The dynamic random effects binary probit model was created by $y_{it}^{b} = 1\left(y_{it}^{b*} > 0\right)$ and $y_{it}^{b} = 0\left(y_{it}^{b*} \leq 0\right)$. The dynamic random effects ordered probit model was created by $y_{it}^{o} = k\left(\tau_{k-1} < y_{it}^{o*} \leq \tau_{k}\right)$. The zero inflation was created by $y_{it} = y_{it}^{b} y_{it}^{o}$ as proposed by Harris and Zhao (2007). The number of Gauss Hermite quadrature nodes and weights was 10 in the experiment. Lesaffre and Spiessens (2001) pointed out that a number of 10 nodes and weights is often adequate and further increment only produces negligible differences. When the algorithm failed to converge, different starting values were utilized to achieve maximisation of the log-likelihood function.

The following parameters were used for simulation. The parameters for the binary model were fixed at $\phi_1 = 0.5$, $\beta_{11} = -1.0$, $\gamma_{11} = -1.0$ and $\sigma_1 = 1.0$. The inter-period correlation was $\sigma_1^2/\left(\sigma_1^2+1\right) = 0.5$. The parameters for the ordinal model were fixed at $\phi_2 = 0.5$, $\gamma_{21} = 0.4$, $\gamma_{22} = 1.0$, $\beta_{21} = 1.0$ and $\sigma_2 = 1.0$. The inter-period correlation was $\sigma_2^2/\left(\sigma_2^2+1\right) = 0.5$. The total number of the replications (R) was 200. The covariance matrix and the correlation matrix were given by .

$$\Sigma_{\delta_1\delta_2} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

$$\Sigma_{ebeo} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

The maximization of the log-likelihood function was done using Newton-Raphson algorithm. This algorithm requires starting values. The starting values for realizing the global maximum of the log-likelihood function were obtained from the dynamic panel binary probit model and DPOP model using the pglm function in the pglm R-package. The study used the estimates from these Dynamic panel binary probit model and DPOP model as the starting values for the Newton-Raphson algorithm. The convergence was speedy based on these estimates. The algorithm was assumed to have converged when the log-likelihood shifted by a small constant, that is, $\varepsilon < 10^{-5}$. The study computed the bivariate normal integrals using pbinorm function in the VGAM R-package.

6.3 Assessing the Performance of the Models

RMSE was used to evaluate the performance the ZIDPOPC and ZIDPOPI models. Let θ be the true parameter. Let $\hat{\theta}_k^r$ be the estimates of k^{th} parameter from r^{th} replication, R be the total number of the replications, and θ_k be the actual value of the k^{th} parameter from the model. The mean of the estimated k^{th} parameter was calculated as follows:

$$\overline{\theta}_k = \frac{1}{R}\sum_{r=1}^{R}\hat{\theta}_k^r \tag{6.1}$$

Root Mean Square Error that includes bias and variability was utilized to assess the accuracy of the estimates. It was calculated for every parameter. It was given by

$$RMSE\left(\theta_k\right) = \sqrt{\frac{\sum_{r=1}^{R}\left(\hat{\theta}_k^r - \theta\right)^2}{R}} \tag{6.2}$$

RMSE is always positive. A value of zero implies a perfect fit to the data. This value is rarely attained in practice. An estimator with a lower value is always preferred than one with a higher value. A lower value indicate a better fit to the data.

6.4 Results for Maximum likelihood estimates based on n=750, T=10 and 10-Point Gauss Hermite quadrature for DPOP, ZIDPOPI and ZIDPOPC models from the simulated data

Table 6.1: Maximum likelihood estimates based on n=750, T=10 and 10-Point Gauss Hermite quadrature for DPOP, ZIDPOPI and ZIDPOPC models from the 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	0.000	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	0.923	-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_{2i2}^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
ρ_{ebeo}	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	

Table 6.1 shows the parameters, their true values, estimates, standard errors, p values and AIC values for DPOP, ZIDPOPI and ZIDPOPC models. All the parameters whose p values were less than 0.05 were significnat at 5%. All the parameters whose p values were less than 0.01 were significant at 1%. 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 associated with individual effect, as evaluated by the intra-unit correlation coefficient in smoking decision. The variance of the individual effects for the decision on consumption levels was 0.674. This indicated that 31.24% of the latent error variance is associated with individual effect, as evaluated by the intra-unit correlation coefficient at consumption levels.

The Akaike Information Criteria (AIC) indicated an assessment of goodness-of-fit. The model with the smallest AIC was considered to provide a better fit for the data than the rest. The ZIDPOPC model clearly fitted the data better than ZIDPOPI and DPOP models. The ZIDPOPI model clearly provided a better fit for the data than DPOP model.

6.5 Assessing the Accuracy of the Estimators in the Models

The accuracy of the three models was evaluated by RMSE.

Table 6.2: Comparison of DPOP, ZIDPOPI and ZIDPOPC when n=350 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			0.542	0.106	0.612	0.151
γ_{11}	-1.0			-0.953	0.066	-0.975	0.045
β_1	-1.0			-0.745	0.314	-0.714	0.299
σ_1	1.0			0.780	0.237	0.810	0.241
ϕ_2	0.5	0.236	0.267	0.544	0.056	0.566	0.085
γ_{21}	0.4	0.157	0.244	0.416	0.037	0.369	0.044
γ_{22}	1.0	0.390	0.610	1.001	0.037	0.972	0.040
β_2	0.7	-0.480	1.180	0.586	0.161	0.387	0.331
σ_2	1.0	0.933	0.101	0.730	0.280	0.676	0.344
ρ_{ebeo}	0.5					0.476	0.072
$\rho_{\delta_{1i}\delta_{2i}}$	0.5			0.264	0.258	0.391	0.208
τ_1	2.6	1.110	1.490	2.567	0.081	2.558	0.114
τ_2	4.2	2.059	2.143	4.150	0.105	4.138	0.111

Table 6.2 shows the parameters, their true values, estimates and RMSE for the three modes when n=350 and T=10.

Table 6.3: Comparison of DPOP, ZIDPOPI and ZIDPOPC when n=750 and T=10 based 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 6.3 shows the parameters, their true values, estimates and RMSE for the three models when n=750 and T=10. for the comparison purpose, Although the ZIDPOPI and ZIDPOPC models had more parameters than DPOP model, for the intent of appraisal, comments were confined to only shared parameters, that is, ϕ_2, γ_{21}, γ_{22}, β_2, σ_2, τ_1 and τ_2. The specific RMSEs of all of the parameters (6 out of 7 compared), were lower in the ZIDPOPI and ZIDPOPC models than the DPOP model for different values of n. This indicated that the ZIDPOPI and ZIDPOPC models' estimates were more accurate than DPOP models' parameters.

6.6 Assessing the Consistency of the Estimators in the Models

Table 6.4: Maximum likelihood estimates based on n=350 and T=10 for DPOP, ZIDPOPI and ZIDPOPC models

MODELS		DPOP		ZIDPOPI		ZIDPOPC	
Par	TRUE	Ests	stderror	Ests	stderror	Ests	stderror
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	0.043	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 6.4 shows the parameters, estimates and their standard errors when n=350 and T=10.

Table 6.5: Maximum likelihood estimates based on n=750 and T=10 for DPOP, ZIDPOPI and ZIDPOPC models

MODELS		DPOP		ZIDPOP		ZIDPOPC	
Par	TRUE	Ests	stderror	Ests	stderror	Ests	stderror
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
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

Table 6.5 shows the parameters, estimates and their standard errors when n=750 and T=10.

The study compared the estimates in Table 5.4 when n=350 and T=10 and Table 6.5 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 example, 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 produced consistent estimators.

6.7 Results for Bayesian Approach from the Simulated Data

The MCMC simulation was run for 40,000 iterations. The first 10,000 iterations were disposed of as they represented the burn-in period. The thinning interval was six. The convergences were analyzed using trace, density and autocorrelation plots. The acceptance rates was maintained between 20% and 30% for all the parameters except the cut points. The acceptance rates was maintained between 25% and 50% for the cut points only as proposed by Cowles (1996).

Table 6.6: Bayesian inference based on n=750, T=10 for DPOP, ZIDPOPI and ZIDPOPC models

MODELS		DPOP		ZIDPOP		ZIDPOPC	
Par	TRUE	Ests	SD	Ests	SD	Ests	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_{2i2}^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
ρ_{ebeo}	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 6.6 indicate the parameters, estimates, standard errors and DIC values. There is a moderate association 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 association between the unobserved heterogeneities 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 was associated with the unobserved heterogeneity, 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 was associated with the unobserved heterogeneity, as evaluated by the intra-unit correlation coefficient at consumption levels. The Deviance Information Criteria (DIC) indicated an assessment of the goodness-of-fit. The model with the smallest DIC is considered to fit the better than the rest. The ZIDPOPC model clearly provided a better than ZIDPOPI and DPOP models. The ZIDPOPI model clearly provided a better than the DPOP model.

6.8 Assessing the Accuracy of the Estimators in the Models

Table 6.7: Bayesian inference based on n=350 and T=10 for DPOP, ZIDPOPI and ZIDPOPC models

MODELS		DPOP		ZIDPOP		ZIDPOPC	
Par	TRUE	Ests	RMSE	Ests	RMSE	Ests	RMSE
ϕ_1	0.5			0.482	0.101	0.532	0.070
γ_{11}	-1.0			-1.053	0.085	-1.056	0.064
β_1	-1.0			-0.757	0.273	-0.779	0.232
σ_1	1.0			0.933	0.070	0.943	0.062
ϕ_2	0.5	0.201	0.300	0.511	0.050	0.540	0.053
γ_{21}	0.4	0.172	0.228	0.406	0.035	0.404	0.033
γ_{22}	1.0	0.419	0.582	1.061	0.087	1.078	0.086
β_2	0.7	-0.490	1.193	0.594	0.199	0.447	0.285
σ_2	1.0	0.861	0.140	0.916	0.085	0.926	0.075
ρ_{ebeo}	0.5					0.430	0.079
$\rho_{\delta_{1i}\delta_{2i}}$	0.5			0.415	0.087	0.433	0.070

τ_1	2.6	1.130	1.472	2.718	0.202	2.787	0.206
τ_2	4.2	2.120	2.082	4.453	0.217	4.511	0.350

Table 6.7 shows the parameters, estimates and RMSE for the three modes when n=350 and T=10.

Table 6.8: Bayesian inference based on n=750 and T=10 for DPOP, ZIDPOPI and ZIDPOPC models

MODELS		DPOP		ZIDPOP		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
ρ_{ebeo}	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

Table 6.8 shows the parameters, estimates and RMSE for the three modes when n=750 and T=10.

Although the ZIDPOPI and ZIDPOPC models had more parameters than DPOP model, for the intent of appraisal, comments are confined to only the shared parameters, that is, ϕ_2, γ_{21}, γ_{22}, β_2, σ_2, τ_1 and τ_2. The majority of the RMSE for the ZIDPOPI and ZIDPOPC models were lower than the DPOP model. The specific RMSE of all of the parameters (6 out of 7 compared), were lower in the ZIDPOPI and ZIDPOPC models than the DPOP model for different values of n. This indicated that the ZIDPOPI and ZIDPOPC models were more accurate than DPOP model.

6.9 Assessing the Consistency of the Estimators in the Models

Table 6.9: Bayesian inference based on n=350 and T=10 for DPOP, ZIDPOPI and ZIDPOPC models

MODELS		DPOP		ZIDPOP		ZIDPOPC	
Par	TRUE	Ests	stderror	Ests	stderror	Ests	stderror
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
ρ_{ebeo}	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	0.1291
τ_2	4.2	2.120	0.062	4.453	0.169	4.511	0.175

Table 6.9 shows the parameter, estimates and thier standard errors for n=350 and T=10.

Table 6.10: Bayesian inference based on n=750 and T=10 for DPOP, ZIDPOPI and ZIDPOPC models

MODELS		DPOP		ZIDPOP		ZIDPOPC	
Par	TRUE	Ests	stderror	Ests	stderror	Ests	stderror
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
ρ_{ebeo}	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

Table 6.10 shows the parameter, estimates and their standard errors for n=750 and T=10.

The study compared estimates in Table 6.9 when n=350 and T=10 and Table 5.10 when n=750 and T=10. The results from Table 6.9 and Table 6.10 indicated that as n increases from 350 to 750, the estimates tend to their true values. For example, the true value of ϕ_1 is 0.5, when n=350, ϕ_1 is 0.532 for ZIDPOPC model and when n=750, ϕ_1 is 0.517. This showed that as n is raised, the estimates tend to the true value. This indicated that ZIDPOPC model produced consistent estimators.

6.10 Application of the Models to Real Life Data

The main dataset used in this thesis was the National Longitudinal Survey of Youth 1997 (NLSY97) that follows the lives of a sample of American youth. The survey is sponsored and directed by the U.S Bureau of Labour Statistics and managed by the Center for Human Resource Research at the Ohio State University. Interviews are conducted by the National Opinion Research Center at the Chicago University. The survey include data on the youths' family and community backgrounds to help researcher assess the impact of schooling and other environmental factors on these labour market entrants. The panel dataset comprised of 8,984 youth who were first interrogated in 1997 when they were between twelve and sixteen years old as of 31 December 1996. Therefore, they were born between 1981 and 1985.

The survey design of the NLSY97 included the smoking habits, socioeconomic and respondent's traits such as race, gender etc. and other variables that are not associated with this study. The study discarded the respondents that had missing values in either the response variable or the covariates. The proportion of missing values were quite high in the early grades as several respondents were let off from the smoking questions in the early stages of the study due to legal and privacy concerns. The study only considered the cohort that was in Grade 6 and above in 1998 and was enrolled in school. The year 1998 was considered the initial observation of the panel data. The study covered eight years of the observations.

The study used the smoking profile used by Subair (2018) in his thesis entitled "Excess Zeros, Endogenous Binary Indicators, and Self-Selection Bias with Application to First Marriage, Smoking and Drinking Outcomes". The study created a smoking profile of 2,500 individuals. The smoking habits of the individuals was based on three questions. The study generated none consumption based on the following questions: "Have you ever smoked a cigarette?" and "During the last 30 days, how many days did you smoke a cigarette?". Individuals who always answered "No" to the first question have absolute inelastic demand for cigarettes, and for such individuals, zero demands for cigarettes are optimal choices. Individuals who answer "None" to the second question are affiliated two categories: those whose optimal choice of zero consumption are defined by corner solutions and infrequent smokers. The study then constructed non-zero consumption from the question "When you smoked during the last 30 days, how many cigarettes did you usually smoke each day?"

Let y_{it} denote the number of sticks of cigarette consumed by a respondent i at time t. y_{it} is the number of sticks of cigarettes smoked on days that respondent i smokes, including none consumption. For instance, in our case, $i = 1, 2, ..., 2500$ and $t = 0, 1, 2, ..., 8$.

Describing the frequency distribution of the crude y_{it} measurements does not actually aid in condensing the smoking intensities for an 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. Therefore, the study created four ordinal outcomes of smoking intensities from their corresponding observed values y_{it} . This can be summarized as follows;

$$y_{it} = \begin{cases} 0 & \text{if respondent } i \text{ is not a current smoker or has never smoked,} \\ 1 & \text{if respondent } i \text{ smokes weekly or less,} \\ 2 & \text{if respondent } i \text{ smokes daily and smokes} < 20 \text{ sticks} \\ 3 & \text{if respondent } i \text{ smokes daily and smokes } \geq \text{ 20 sticks.} \end{cases} \tag{6.3}$$

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

6.11 Frequency Distribution Tables of Real Data

This section present the summarised statistics of smoking intensities and all their explanatory variables.

Table 6.11: Distribution of Smoking Intensities 1998-2001

Ordinal	1998		1999		2000		2001	
Outcomes	N	%	N	%	N	%	N	%
0	1889	75.56	1776	71.04	1700	68.00	1640	65.60
1	386	15.44	444	17.60	456	18.24	505	20.20
2	167	6.68	199	7.96	221	8.84	224	8.96
3	58	2.52	85	3.40	123	4.92	131	5.24

The Table 6.11 shows the frequency and percentages of ordinal outcomes between 1998 and 2001 for the National Longitudinal Survey of Youth 1997 (NLSY97).

Table 6.12: Distribution of Smoking Intensities 2002-2004

Ordinal	2002		2003		2004		2005		2006	
Outcomes	N	%	N	%	N	%	N	%	N	%
0	1592	63.68	1582	63.28	1560	62.40	1585	63.40	1594	63.76
1	517	20.68	486	19.44	506	20.24	489	19.56	469	18.76
2	239	9.56	276	11.04	258	10.32	251	10.04	254	10.16
3	152	6.08	156	6.24	176	7.04	175	7.00	183	7.32

The Table 6.12 shows the frequency and percentages of ordinal outcomes between 2002 and 2006 for the National Longitudinal Survey of Youth 1997 (NLSY97).

Table 6.13: Distribution of Smoking status since the last date of interview 1998-2002

Ordinal	1998		1999		2000		2001		2002	
Outcomes	N	%	N	%	N	%	N	%	N	%
No smoke	1602	64.08	1586	63.44	1524	60.96	1492	59.68	1443	57.72
Smoke	898	35.92	914	36.56	976	39.04	1008	40.32	1057	42.28

The Table 6.13 shows the frequency and percentages of binary outcomes between 1998 and 2002 of whether the respondent smoke or not since the last day of interview for the National Longitudinal Survey of Youth 1997 (NLSY97).

Table 6.14: Distribution of Smoking status since the last date of interview 2003-2006

Ordinal	2003		2004		2005		2006	
Outcomes	N	%	N	%	N	%	N	%
No smoke	1432	57.28	1450	58.00	1458	58.32	1464	58.56
Smoke	1068	42.72	1050	42.00	1042	41.68	1036	41.44

The Table 6.14 shows the frequency and percentages of binary outcomes between 2003 and 2006 of whether the respondent smoke or not since the last day of interview for the National Longitudinal Survey of Youth 1997 (NLSY97).

Table 6.15: Current Age of the Respondent 1998-2006

Years	Min	Median	Mean	Max
1998	13.00	16.00	15.83	19.00
1999	14.00	17.00	16.80	20.00
2000	15.00	18.00	17.88	21.00
2001	16.00	19.00	18.87	22.00
2002	18.00	20.00	19.87	23.00
2003	18.00	21.00	20.82	24.00
2004	19.00	22.00	21.86	25.00
2005	20.00	23.00	22.83	26.00
2006	21.00	24.00	23.78	27.00

The Table 6.15 shows the summary of age between 1998 and 2006 for the National Longitudinal Survey of Youth 1997 (NLSY97).

Table 6.16: Distribution of Gender

Gender	N	%
Male	1233	49.32
Female	1267	50.68

The Table 6.16 shows the frequency and percentages of gender between 1998 and 2006 for the National Longitudinal Survey of Youth 1997 (NLSY97).

Table 6.17: Distribution of Race

Race	N	%
White	2109	84.36
Black	391	15.64

The Table 6.17 shows the frequency and percentages of race between 1998 and 2006 for the National Longitudinal Survey of Youth 1997 (NLSY97).

6.12 Results for the Maximum Likelihood Estimation from the Smoking Data

Table 6.18: Maximum likelihood estimates based on n=2500 and T=8 for DPOP, ZIDPOPI and ZIDPOPC Models from the Smoking data

MODELS	DPOP			ZIDPOP			ZIDPOPC		
Var	Ests	Stderr	p value	Ests	Stderr	p value	Ests	Stderr	p value
ϕ_1				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
interc	-1.477	0.320	0.000	-0.687	0.389	0.078	0.602	0.517	0.244
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^o	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
ρ_{ebeo}							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 6.18 shows the variables, estimates, standard errors and p values. The variables whose estimates are positive means an increase in the variable leads to in increase in the predicted probability of smoking and verse versa. For instance, the estimate of age in ZIDPOPC model was 0.026 implying an increase in age will results into an increase in

the predicted probability of smoking. The variables whose p values are less than 0.05 are significant at 5%. In case of ZIDPOPC model, state depedence, race and initial observation from the decision to smoke or not are are signficant at 5% while state dependence, age, gender, initial observation, correlation between individual effects and the first cut point from decision on consumption level 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 was associated with the individual effects, as evaluated by the intra-unit correlation coefficient in smoking decision. The variance of the individual effects for the decision on consumption levels was 0.954. This indicated that 47.65% of the latent error variance was associated with the individual effects , as evaluated 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 obtained coefficients for first period in smoking decision was significant at 1%, which indicated a positive association between the first period smoking decision and latent smoking variable. Hence, this signify the importance of controlling for smoking decision at the beginning of observations. The obtained coefficients for first period decision on the number of cigarettes smoked observation was significant at 1%, which implied a positive association between the first period consumption levels observation and latent consumption levels. Hence, this signify the importance of controlling the reported smoking intensities at the beginning of the observations.

The Akaike Information Criteria (AIC) indicated an assessment of goodness-of-fit. The model with the smallest AIC is considered to fit the data better than the rest. The ZIDPOPC model clearly provided a better fit than the ZIDPOPI and DPOP models. The ZIDPOPI model is clearly provided a better than the DPOP models.

Table 6.19: Average Partial Effects for variables in ZIDPOPC model from the Real data

Var		Ests	Stderror	t values	p value
ϕ_1	$y_{it}^b = 0$	0.328	0.031	10.524	0.000
Age	$y_{it}^b = 0$	-0.003	0.000	-30.423	0.000
Edu	$y_{it}^b = 0$	-0.007	0.000	-32.041	0.000
Gender	$y_{it}^b = 0$	0.018	0.001	14.633	0.000
Race	$y_{it}^b = 0$	-0.057	0.005	-10.941	0.000
ϕ_1	$y_{it}^o = 0$	-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 6.19 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 was 0.328. This indicated that as one move from the previous year to the next, the probability of smoking would increase by 32.8%. The average partial effects for the age in non-participation decision was -0.003. This indicated that single-unit rise in age will provides a 0.3% decrease in the chances of smoking for an otherwise "average" individual while. The average partial effects for the education in non-participation decision was -0.007. This indicated that single-unit rise in age would provide a 0.7% decrease in the chances of smoking for an otherwise "average" individual. The average partial effects for the gender in non-participation decision was 0.018. This indicated that as one change from female to male, the probability of smoking and smoking would increase by 1.8%. The average partial effects for the gender in non-participation decision was -0.057. This indicated that as one change from white to black race, the probability of smoking would decrease by 5.7%.

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 would 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 education, at various consumption levels, that is, zero, low, moderate and high consumption levels were -0.005, 0.007, 0.004 and 0.004, respectively. This indicated that one-unit increase in education would produce -0.5%, 0.7%, 0.4% and 0.4% change in various consumption levels, that is , zero, low, moderate and high consumption levels respectively

The average partial effects for the gender, at various consumption levels, that is, zero, low, moderate and high consumption levels were 0.007, 0.436, 0.194 and 0.024, 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 0.7%, 43.6%, 19.4% and 2.4% 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.

6.13 Results for Bayesian inference for Real Data

Table 6.20: Bayesian inference based on n=2500 and T=8 for DPOP, ZIDPOPI and ZIDPOPC Models from the Smoking data

MODELS	DPOP		ZIDPOP		ZIDPOPC	
Var	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
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_2^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
ρ_{ebeo}					0.116	0.080
$\rho_{\delta_{1i}\delta_{2i}}$			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 6.20 shows estimated parameters and their standard errors for the three models. The variables whose estimates are positive means an increase in the variable leads to in increase in the predicted probability of smoking and verse versa. For instance, the estimate of age in ZIDPOPC model was -0.005 implying an increase in age will results into a decrease in the predicted probability of smoking. The variance of individual effects for participation was 0.986. This indicated that 49.30% of the latent error variance was associated with the individual effects, as evaluated by the intra-unit correlation coefficient in smoking decision. The variance of individual effects for consumption levels was 0.976. Approximately 48.79% of the latent error variance was associated with the individual effects, 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 Deviance information criterion (DIC) indicated an assessment of goodness-of-fit. The model with the smallest DIC was considered to provide a better than the rest. The ZIDPOPI model clearly provided a 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 provided a better than the DPOP model.

Table 6.21: Average Partial Effects for variables in Correlated Zero Inflated Dynamic Panel Ordered Probit Model (ZIDPOPC) for Real data

Var		Ests	Stderror	t value	p value
ϕ_1	$y_{it}^b = 0$	0.353	0.026	13.465	0.000
Age	$y_{it}^b = 0$	0.001	0.000	26.737	0.000
Edu	$y_{it}^b = 0$	-0.012	0.001	-20.079	0.000
Gender	$y_{it}^b = 0$	0.019	0.001	14.014	0.000
Race	$y_{it}^b = 0$	-0.055	0.005	-10.503	0.000
ϕ_1	$y_{it}^0 = 0$	-0.055	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_2	$y_{it}^0 = 0$	-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 6.21 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 indicated a positive association with the participation and consumption at all levels and vice versa.

The average partial effects for the state dependence in non-participation decision and participation decision were 0.353. This indicated that as one move from the previous year to the next, the probability of smoking would increase by 35.3%. The average partial effects for the age in non-participation decision was 0.001. This indicated that single-unit rise in age will produce a 0.1% increase in the chances of smoking for an otherwise "average" individual. The average partial effects for the education in non-participation decision was -0.012. This indicated that one-unit increase in age would produce a -1.2% decrease in the probability of smoking for an otherwise "average" individual. The average partial effects for the gender in non-participation decision and participation decision was 0.019. This indicated that as one change from female to male, the probability of smoking would increase by 1.9%. The average partial effects for the gender in non-participation decision was -0.055 and -0.061. This indicated that as one change from white to black race, the probability of smoking would decrease by 5.5%.

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 would 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 education, at various consumption levels, that is, zero, low, moderate and high consumption levels were -0.006, 0.011, 0.004 and 0.004, respectively. This indicated that one-unit increase in education would produce -0.6%, 1.1%, 0.4% and 0.4% change in various consumption levels, that is, zero, low, moderate and high consumption levels respectively. The average partial effects for the gender, at various consumption levels, that is , zero, low, moderate and high consumption levels were 0.008, 0.419, 0.196 and 0.022 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 0.8%, 41.9%, 19.6% and 2.2% 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

CHAPTER SEVEN

SUMMARY, CONCLUSION AND RECOMMENDATIONS

7.1 Introduction

This chapter discusses the summary, conclusions, recommendation and suggestions for further research.

7.2 Summary

The purpose of this thesis was to develop a Zero inflated dynamic panel ordered probit model with both independent and uncorrelated error terms and compare them with Dynamic panel ordered probit model. Unlike the Semi-parametric Zero inflated dynamic panel probit model with selectivity proposed by Christelis and Galdeano (2009) that assumed that the unobserved individual effects have a non-parametric distribution, the develop Zero inflated dynamic panel ordered probit model with both independent and uncorrelated error terms in Chapter Three in this thesis are considered to follow a bivariate normal distribution that has a parametric distribution.

The parametric distribution enabled the estimation of variances in both participation and consumption levels that facilitated the estimation of inter-unit correlation and correlation between the unobserved individual effects. The inter-unit correlation facilitated the determination of latent error variance that was associated with the individual effects. The correlation between the unobserved individual effects explained whether the factors affecting unobserved individual effects in participation decision were the same as those affecting the unobserved individual effects at consumption levels. The heaping in the Semi-parametric Zero inflated dynamic panel probit model proposed by Christelis and Galdeano (2009) allowed the choice of cut points to address bunching in the number of cigarettes smoked and identify the variance of the dynamic panel ordered probit model. The Zero inflated dynamic panel ordered probit model with both correlated and uncorrelated error terms developed in this thesis were assumed to contain unknown cut points due to absence of heaping.

Instead of using Akay (2009) approach for initial conditions that was found to yield severe bias in estimation by Rabe-Hesketh and Skrondal (2013), the study used their approach that used first response variable, initial covariates and within-means of covariates omitting first values. The Semi-parametric Zero inflated dynamic panel probit model proposed by Christelis and Galdeano (2009) used simulated maximum likelihood approach proposed by Lee and Oguzoglu (2007) and Kano (2008) where the individual's effects are integrated out by computing the double integral by simulation. Mulkay (2015) pointed that this procedure could be very time-consuming even with fast modern computer. This thesis used an alternative approach in Chapter Three proposed by Raymond *et al.* (2007) based on a two-step Gauss-Hermite Quadrature to compute the double integral. The two-step Gauss-Hermite Quadrature have a satisfactory performance in finite sample for a small number of nodes and weights. The performance of the models was evaluated using Akaike Information Criteria. The model with smallest AIC was considered to fit data better than the rest.

The two-step Gauss-Hermite Quadrature can results into high computational cost. The high cost of direct computing the double integrals can be reduced in a Bayesian approach. This study proposed Bayesian approach in Chapter Four for the Zero inflated dynamic panel ordered probit model with both independent and correlated error terms using Metropolis Hasting and Gibb sampling and evaluated the convergence using autocorrelation, density and trace plots. The performance of the models was evaluated using Deviance Information Criteria. The model with smallest DIC was considered to provide a better fit than the rest. The Average partial effects for the Zero inflated dynamic panel ordered probit model with both independent and correlated error terms were also presented in Chapter Four to facilitate the interpretations of the results.

Simulation studies and discussions were presented in Chapter Five. The Simulation studies were used to evaluate the theoretical properties of DPOP, ZIDPOPI and ZIDPOPC estimators. The theoretical properties are the accuracy and consistency of the estimators in the model. The chapter also contain an empirical study where the models were used to determine the persistence in decision to participate in smoking and the decision on consumption levels using the National Longitudinal Survey of

Youth 1997 (NLSY97) from American youths born between 1980-1984.

7.3 Conclusion

The main objective of this thesis was to develop a Zero inflated dynamic panel ordered probit models with both correlated and uncorrelated error terms and compare them with Dynamic panel ordered probit model. The Akaike information criteria and deviance information criteria from the simulation studies showed that ZIDPOPC model provided a better than ZIDPOPI and DPOP models while ZIDPOPI model provided a better than DPOP model. This results are similar to Harris and Zhao (2007) and Gurmu and Dagne (2009) based on non-dynamic versions of the same models. ZIDPOPI and ZIDPOPC estimators had a smaller RMSE compared to DPOP estimators for different values of n. This showed that ZIDPOPC and ZIDPOPI estimators were more accurate than DPOP estimators. The estimates of the ZIDPOPI and ZIDPOPC models tended to the parameter values as n tend to infinity. This indicated that ZIDPOPI and ZIDPOPC models produced consistent estimators. 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. Christelis and Galdeano (2009) also found a statistically significant and economically relevant estimate of true state dependence in most European countries.

The existence of state dependence implies that short-term policy interventions established to reduce participation and consumption levels may have longer-term implications. The estimated parameter for the first period participation observations were statistically significant at 1%, which implied a positive association between the first period participation observation and latent participation observation. Similarly, the estimated parameter for first period consumption observations were statistically significant at 1%, which implied a positive association between the first period consumption observation and latent consumption observation. This implied that it is essential to control for participation decision and consumption levels at the beginning of the observations.

For classical approach, 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. The variance of the individual effects in participation decision was 0.771. This indicated that 37.28% of the latent error variance was associated with the individual effects, as evaluated by the intra-unit correlation coefficient in smoking decision. The variance of the individual effects for the decision on consumption levels was 0.674. This indicated that 31.24% of the latent error variance was associated with the individual effects, as evaluated by the intra-unit correlation coefficient at consumption levels.

For Bayesian approach, the variance of the individual effects in participation decision was 0.938. This indicated that 46.80% of the latent error variance was associated with the individual effects, as evaluated by the intra-unit correlation coefficient in smoking 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 was associated with the individual effects, as evaluated by the intra-unit correlation coefficient at consumption levels. There was a moderate correlation (0.470) between the unobserved individual effects in ZIDPOPC model, this implied that the factors affecting the individual effects in participation decision were same as the one affecting the unobserved individual effects at consumption levels.

The Akaike information criteria from the smoking data in maximum likelihood estimation showed that ZIDPOPC model provided a better than ZIDPOPI and DPOP models while ZIDPOPI model provided a better than DPOP model. In case of ZIDPOPC model, state dependence, race and initial observation from the participation decision were significant at 5% while state dependence, age, gender, initial observation, correlation between individual effects and the first cut point from consumption level were significant at 5%. The main causes of persistence of decision to smoke were state dependence, unobserved heterogeneity, initial observations and race. Age, education and gender do not influence the decision to smoke or not among the young youth. This is similar to Contoyannis *et al.* (2004) who pointed out that the explanatory power of majority of the covariates vanishes if state dependence and individual effects are controlled for. Therefore, state dependence and individual effects should be a major focus to policymakers who have a smoking participation as an objective.

The main causes of persistence for consumption levels were state dependence, unobserved heterogeneity, age and gender. Educational level do not influence the consumption levels among the young youth. On average, age has a negative impact at zero consumption but a positive impact, particularly at higher percentiles, on probability of cigarette consumption as the intensity of smoking increases. This results are similar to Gurmu and Dagne (2009). This may be attributed to addictive behavior as one gets older. The impacts in Gurmu and Dagne (2009) are overestimated due to overlooking of state dependence in their model. Again, this is similar to Contoyannis *et al.* (2004) who pointed out that the explanatory power of majority of the covariates vanish if state dependence and individual effects are controlled for. Therefore, state dependence and individual effects should be a major focus to policymakers who have a smoking participation as an objective.

The state dependence in participation and consumption levels of cigarettes are characterized by substantial positive state dependence. This implied that short-term policy interventions designed to reduce participation in smoking and consumption levels of cigarettes may have longer-term implications.The variance of individaual effects in participation was 0.998. This indicated that 49.90% of the latent error variance is associated with unobserved heterogeneity, as evaluated by the intra-unit correlation coefficient in smoking decision. The variance of individual effects for consumption levels was 0.954. Approximately 47.65% of the latent error variance is associated with unobserved heterogeneity, as evaluated by the intra-unit correlation coefficient at consumption levels.

The correlation coefficient between the errors was not signficant at 5%. This implied that the variables affecting the participation 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 were the same as the one affecting individual effects at consumption levels. Estimated coefficients for initial period smoking observations were significant at 1%, for both smoking decision and consumption levels, which implied a positive correlation between the initial period smoking observation and unobserved latent smoking. Therefore, this indicates that it is essential to control for the reported smoking at the beginning of the observations.

The estimated parameters for first period smoking decision was significant at 1%, which implies a positive association between the first period smoking observation and latent smoking observation. Hence, it is important to control for smoking decision at the beginning of observations. The estimated parameters for first 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 first period consumption levels observation and latent consumption levels. Hence, it is important to control for the reported smoking at the beginning of the observations.

The Deviance information criteria from the smoking data in Bayesian approach showed that ZIDPOPI model provided a better than ZIDPOPC and DPOP models while ZIDPOPC model provided a better than DPOP model. This is due to the weak correlation between the error terms in participation and consumption levels of cigarettes. The state dependence in participation and consumption levels of cigarettes are depicted by considerable positive state dependence. This implied that short-term policy interventions designed to reduce participation in smoking and consumption levels of cigarettes may have longer-term implications.

The variance of individual effects for participation was 0.986. This indicated that 49.30% of the latent error variance was associated with the individual effects, 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 was associated with the individual effects, 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.

Since the Zero inflated dynamic panel ordered probit model with independent and correlated error terms provided a better than Dynamic panel ordered probit model in both classical and Bayesian approaches, they are recommended for use in practice.

7.4 Recommendations for Future Research

This study considered the Zero inflated dynamic panel ordered probit model in a balanced panel model and in absence of attrition. In real life, some respondent may drop out of the study due to a variety of reason such as death or relocating to a different country. Some respondent may not be available at the beginning of the study but are likely to join the study later. This results to an unbalanced panel model where some respondent are not observed for all the period. This study can be extended to an unbalanced panel model and assume presence of attrition. The models can also be extended to handle missing values.

REFERENCES

Abramovitz, M., and Stegun, I. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Washington: National Bureau of Standards Applied Mathematics, US Government Printing Office.

Acero, J. A. P., and Luis, J. J. (2019). Persistence in the imitation of innovations in products in the manufacturing industry of colombia. *Especial Innovación*, *64(1)*, 1-15.

Aitchison, J., and Silvey, S. D. (1957). The generalization of probit analysis to the case of multiple reponses. *Biometrika*, *44*, 131–140.

Akay, A. (2009). Finite-sample comparison of alternative methods for estimating dynamic panel data models. *Journal of Applied Econometrics*, *27*, 1189–1204.

Albert, J. H., and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, *44*, 669-679.

Arulampalam, W. (1999). A note on estimated coefficients in random effects probit models. *Oxford Bulletin of Economics and Statistics*, *61*, 597-602.

Ayllon, S., and Blanco, P. (2012). State dependence in self-assessed health in spain. *Hacienda Pública Española*, *202*, 9-30.

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.

Bartholomew, D. J. (1983). Latent variable models for ordered categorical data. *Journal of Econometrics*, *22*, 229-243.

Beck, N., Epstein, D., and Jackman, S. J. (2012). Alternative models of dynamics in binary time-series–cross-section models: The example of state failure. In *2001 annual meeting of the society for political methodology.* Emory University.

Bliss, C. I. (1934). The method of probits. *Science*, *79*, 38–39.

Brooks, R., Harris, M. N., and Spencer, C. (2008). An inflated ordered probit model of monetary policy: Evidence from mpc voting data. *Economic Letter*, *117*, 683-686.

Bureau of Labor Statistics United States of America, U. (2019). *National longitudinal survey of youth 1997 cohort, 1997-2017 (rounds 1-18).* Produced and distributed by the Center for Human Resource Research (CHRR). The Ohio State University. Columbus.

Butler, J. S., and Moffitt, R. (1982). A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica*, *50*, 761-764.

Butller, A. M. (2011). Hierarchical probit models for ordinal ratings data. *All Theses and Dissertations*, *2656*.

Caglayan, A. E., and Van, M. H. (2017). Determinants of the levels of development based on the human development index: Bayesian ordered probit model. *International Journal of Economics and Financial Issues*, *7(5)*, 425-431.

Chib, S., and Albert, J. H. (2001). Sequential ordinal modeling with applications to survival data. *Biometrics*, *57*, 829-836.

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.

Contoyannis, P., Jones, A., and Rice, R. (2004). The dynamics of health in the british household panel study. *Journal of Applied Econometrics*, *19*, 473-503.

Cowles, M. K. (1996). Accelerating monte carlo markov chain convergence for cumulative- link generalized linear models. *Statistics and Computing*, *6*, 101–111.

Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica: Journal of the Econometric Society*, 829-844.

Das, S., Hossain, M. E., and Akter, K. (2021). Students' perception of online class during covid-19 pandemic: An ordered probit model estimation. *International Journal of Academic Research in Progressive Education and Development*, *10(2)*, 392–401.

David, D., and Sirchenko, A. (2018). *Estimation of nested and zero-inflated ordered probit models.* (Working paper ECONOMICS WP BRP 193/EC/2018)

Downward, P., Lera-Lopez, F., and Rasciute, S. (2011). The zero-inflated ordered probit approach to modelling sports participation. *Economic Modelling*, *28 (6)*, 2469-2477.

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.

Fielding, A., and Yang, M. (2005). Generalized linear mixed models for ordered responses in complex multilevel structures: effects beneath the school or college in education. *Journal of the Royal Statistical Society Series a-Statistics in Society*, *168*, 159-183.

Franz, P., Raymond, W. P., and Sybrand, S. V. D. (2010). Persistence of innovation in dutch manufacturing: Is it spurious? *Review of Economics and Statistics*, *92(3)*, 495-504.

Gallefoss, E., and Bakke, P. S. (2000). Patient satisfaction with healthcare in asthmatics and patients with copd before and after patient education. *Respiratory Medicine*, *94*, 1057-1064.

Gelfand, A. E., and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*, 398-409.

Geyer, C. J. (1992). Practical markov chain monte carlo (with discussion). *Statistical Science*, *7*, 473—511.

Greene, W., Harris, M. N., Srivastava, P., and Zhao, X. (2018). Misreporting and econometric modelling of zeros in survey data on social bads: An application to cannabis consumption. *Health Economics*, *27*, 372–389.

Guilkey, D. K., and Murphy, J. L. (1993). Estimation and testing in the random effects probit model. *Journal of Econometrics*, *59*, 301-317.

Gurmu, S., and Dagne, A. G. (2012). Bayesian approach to zero-inflated bivariate ordered probit regression model, with an application to tobacco use. *Journal of Probability and Statistics*, 1-26.

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.

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.

Hasegawa, H. (2009). Bayesian dynamic panel-ordered probit model and its application to subjective well-being. *Communications in statistics-simulation and computation*, *38*, 1321-1347.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, *57(1)*, 97-109.

Heckman, J. J. (1981a). *Heterogeneity and state dependence*. Chicago: S. Rosen ed. Studies in labor markets: University of Chicago Press.

Heckman, J. J. (1981b). *The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process and some monte carlo evidence," in c.f. manski and d. mcfadden eds. structural analysis of discrete data with econometric applications*. Cambridge: MIT press.

Jackman, S. (2000a). Estimation and inference via bayesian simulation: An introduction to markov chain monte carlo. *American Journal of Political Science*, *44*, 375–404.

Jackman, S. (2000b). *2000. "in and out of war and peace: Transitional models of* *international conflict." working paper.*

Kano, S. (2008). *Like husband, like wife: A bivariate dynamic probit analysis of spousal obesities.* Osaka Prefecture University.

Kitenge, E. (2020). Covid-19: A virus for the rich and the poor. *SSRN Electronic Journal*.

Kostecki-Dillion, T., Monette, G., and Wong, P. (1999). *Pine trees, comas and migraine* (No. 14(2)).

Lambert, D. (1992). Zero-inflated poisson regression with an application to defects in manufacturing. *Technometrics*, *31*, 1-14.

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.

Lesaffre, E., and Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random-effects model: an example. *Appl. Statist.*, *50*, 325–335.

Liu, I., and Agresti, A. (2005). *The analysis of ordered categorical data: An overview and a survey of recent developments*. Test. Retrieved from `https://doi.org/10.1007/BF02595397`

Martin, J. C., and Concepcion, R. (2021). Covid-19 is examining the eu and the member states: The role of attitudes and sociodemographic factors on citizens' support towards national policies. *Social Sciences 10*, *46*.

McCullagh, P. (1980). Regression models for ordinal data (with discussion). *J Roy Statist Soc*, *42*, 109-142.

McKelvey, R., and Zavoina, W. (1975). A statistical model for the analysis of ordered level dependent variables. *Journal of Mathematical Sociology*, *4*, 103–120.

Metropolis, N., and Ulam, S. (1949). The monte carlo method. *Journal of the American Statistical Association*, *44(247)*, 335-341.

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.* Université de Montpellier, Montpellier.

Mullahy, J. (1997). Heterogeneity, excess zeros and the structure of count data models. *Journal of Applied Econometrics*, *12*, 337-350.

Muller, G., and Czado, C. (2005). An autoregressive ordered probit model with application to high-frequency finance. *Journal of Computational and Graphical Statistics*, *14*, 320–338.

Murphy, D. A., Brecht, M. L., Herbeck, D., Evans, E., Huang, D., and Hser, Y. L. (2008). Longitudinal hiv risk behavior among the drug abuse treatment outcome studies (datos) adult sample. *Evaluation Review*, *32*, 83-112.

Muthen, B., and Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*, 171-189.

Neyman, J., and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Journal of the Econometric Society*, *16*, 1–32.

Oh, M., Park, S., and Oh, H. T. (2012). Bayesian analysis of korean alcohol consumption data using a zero-inflated ordered probit model. *Korean Journal of Applied Statistics*, *25*, 363-376.

Ohh, H. S. (2015). Comparison of waist-to-height ratio (whtr), body mass index (bmi) and waist circumference (wc) as a screening tool for prediction of metabolic-related diseases. *Journal of Chosun Natural Science*, *8*, 305- 312.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*, 443-460.

Park, J. (2012). *Essays on bayesian inference of time-series and ordered panel data models* (PhD thesis). State University of New Jersey.

Plackett, R. L. (1954). A reduction formula for normal multivariate integrals. *Biometrika*, *41*, 351-360.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2005). Output analysis and diagnostics for mcmc. *R package version*, *10-3*.

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.

Rajendra, K. (2013). *A latent mixture approach to modeling zero-inflated bivariate ordinal data* (PhD thesis). University of South Florida.

Raymond, W., Mohnem, P., Palm, F., and Loeff, S. S. D. (2007). *The behavior of the maximum likelihood estimator of dynamic panel data sample selection models* (Tech. Rep. No. 1992).

Seibert, A., Sirchenko, A., and Müller, G. (2018). *A model for policy interest rates.* (Working paper ECONOMICS WP BRP 192/EC/2018)

Spiegelhalter, D. J., Best, N., Carline, B., and Van, D. L. (2002). Bayesian method of model complexity and fit. *Journal of the Royal Statistical Society*, *64*, 583-639.

Stegmueller, D. (2013). Modeling dynamic preferences. a bayesian robust dynamic latent ordered probit model. *Political Analysis*, *21(3)*, 314-333.

Subair, L. A. (2018). *Excess zeros, endogenous binary indicators, and self-selection bias with application to first marriage, smoking and drinking outcomes* (PhD thesis). University of Mississippi.

Swinerton, D. (2021). Probit analysis of the effect of covid-19 job characteristics on unemployment. *Digital Student Showcase Spring*.

Tanner, M. A., and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, *82*, 528-550.

Taylor, J. (1993). Discrete versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy*, *38*, 195-214.

Varin, C., and Czado, C. (2010). Modeling migraine severity with autoregressive ordered probit model. *Biometrics*, *1*, 1–19.

Varin, C., and Vidoni, P. (2006). Pairwise likelihood inference for ordinal categorical time series. *Computational Statistics and Data Analysis*, *51*, 2365-2373.

Williamson, J. M., Kim, K. M., and Lipsitz, S. R. (1995). Analyzing bivariate ordinal data using a global odds ratio. *Journal of the American Statistical Association*, *90*, 1432-1437.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge: MIT press.

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.

Xin-She, Y. (2020). *Mathematical modeling with multidisciplinary applications*. John Wiley and Sons, Inc.

Yong-Woo, L. (2016). State dependence, unobserved heterogeneity, and health dynamics in korea. *Hitotsubashi Journal of Economics*, *57(2)*, 195-221.

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.

Zuur, G., Garthwaite, P. H., and Fryer, R. J. (2002). Practical use of mcmc methods:lessons from a case study. *Biometrical Journal*, *44(4)*, 433-455.

APPENDIX

I Second -Order Differentiations of the Log-Likelihood Function

The MLE for Θ can be availed by equating first derivatives to 0 and solving for Θ. But the expressions can only be solved by an iterative method, for example by the Newton-Raphson iterative method. To apply the Newton-Raphson method we require the second derivatives expressions in the Hessian matrix and also require $\frac{\partial \pi_{ki}}{\partial \gamma_1}$.

$$\begin{aligned}
\frac{\partial \pi_{qhi}}{\partial \gamma_1} &= \frac{\partial}{\partial \gamma_1} \frac{w_q w_h f(y_{itk}|\Theta)}{\sum\limits_{q=1}^{Q}\sum\limits_{h=1}^{H} w_q w_h f(y_{itk}|\Theta)} \\
&= \frac{w_q w_h \frac{\partial f(y_{itk}|\Theta)}{\partial \gamma_1}}{\sum\limits_{q=1}^{Q}\sum\limits_{h=1}^{H} w_q w_h f(y_{itk}|\Theta)} - \frac{w_q w_h f(y_{itk}|\Theta) \sum\limits_{q=1}^{Q}\sum\limits_{h=1}^{H} w_q w_h \frac{\partial f(y_{itk}|\Theta)}{\partial \gamma_1}}{\left(\sum\limits_{q=1}^{Q}\sum\limits_{h=1}^{H} w_q w_h f(y_{itk}|\Theta)\right)^2} \\
&= \frac{w_q w_h f(y_{itk}|\Theta) \frac{\partial \log f(y_{itk}|\Theta)}{\partial \gamma_1}}{\sum\limits_{q=1}^{Q}\sum\limits_{h=1}^{H} w_q w_h f(y_{itk}|\Theta-)} - \\
&\quad \frac{w_q w_h f(y_{itk}|\Theta) \sum\limits_{q=1}^{Q}\sum\limits_{h=1}^{H} w_q w_h f(y_{itk}|\Theta) \frac{\partial \log f(y_{itk}|\Theta)}{\partial \gamma_1}}{\sum\limits_{q=1}^{Q}\sum\limits_{h=1}^{H} w_q w_h f(y_{itk}|\Theta) \sum\limits_{q=1}^{Q}\sum\limits_{h=1}^{H} w_q w_h f(y_{itk}|\Theta)} \\
&= \frac{w_q w_h f(y_{itk}|\Theta) \frac{\partial \log f(y_{itk}|\Theta)}{\partial \gamma_1}}{\sum\limits_{q=1}^{Q}\sum\limits_{h=1}^{H} w_q w_h f(y_{itk}|\Theta)} - \frac{w_q w_h f(y_{itk}|\Theta)}{\sum\limits_{q=1}^{Q}\sum\limits_{h=1}^{H} w_q w_h f(y_{itk}|\Theta)} * \\
&= \frac{\sum\limits_{q=1}^{Q}\sum\limits_{h=1}^{H} w_q w_h f(y_{itk}|\Theta) \frac{\partial \log f(y_{itk}|\Theta)}{\partial \gamma_1}}{\sum\limits_{q=1}^{Q}\sum\limits_{h=1}^{H} w_q w_h f(y_{itk}|\Theta)} \\
&= \pi_{qhi} \frac{\partial \log f(y_{itk}|\Theta)}{\partial \gamma_1} - \pi_{qhi} \sum_{q=1}^{Q}\sum_{h=1}^{H} \pi_{qhi} \frac{\partial \log f(y_{itk}|\Theta)}{\partial \gamma_1}
\end{aligned} \tag{1}$$

The second derivatives are given by,

$$
\begin{aligned}
\frac{\partial^2 \ell(\Theta,y)}{\partial\gamma_1\partial\gamma_1{}'} &= \frac{\partial}{\partial\gamma_1}\left(\sum_{i=1}^{N}\sum_{q=1}^{Q}\sum_{h=1}^{H}\pi_{qhi}\frac{\partial \log f(y_{itk}|\Theta)}{\partial\gamma_1}\right) \\
&= \sum_{i=1}^{N}\sum_{q=1}^{Q}\sum_{h=1}^{H}\left(\pi_{qhi}\frac{\partial^2 \log f(y_{itk}|\Theta)}{\partial\gamma_1\partial\gamma_1{}'} + \frac{\partial \log f(y_{itk}|\Theta)}{\partial\gamma_1}\frac{\partial\pi_{qhi}}{\partial\gamma_1}\right) \\
&= \sum_{i=1}^{N}\sum_{q=1}^{Q}\sum_{h=1}^{H}\left(\pi_{qhi}\frac{\partial^2 \log f(y_{itk}|\Theta)}{\partial\gamma_1\partial\gamma_1{}'} + \frac{\partial \log f(y_{itk}|\Theta)}{\partial\gamma_1}\right. \\
&\quad \left.\left(\pi_{qhi}\frac{\partial \log f(y_{itk}|\Theta)}{\partial\gamma_1} - \pi_{qhi}\sum_{q=1}^{Q}\sum_{h=1}^{H}\pi_{rli}\frac{\partial \log f(y_{itk}|\Theta)}{\partial\gamma_1}\right)\right) \\
&= \sum_{i=1}^{N}\sum_{q=1}^{Q}\sum_{h=1}^{H}\left(\pi_{qhi}\frac{\partial^2 \log f(y_{itk}|\Theta)}{\partial\gamma_1\partial\gamma_1{}'} + \pi_{qhi}\left(\frac{\partial \log f(y_{itk}|\Theta)}{\partial\gamma_1}\right)^2\right. \\
&\quad \left. - \frac{\partial \log f(y_{itk}|\Theta)}{\partial\gamma_1}\pi_{qhi}\sum_{q=1}^{Q}\sum_{h=1}^{H}\pi_{rli}\frac{\partial \log f(y_{itk}|\Theta)}{\partial\gamma_1}\right)
\end{aligned} \tag{2}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell(\Theta,y)}{\partial\gamma_{01}\partial\gamma_1} &= \frac{\partial}{\partial\gamma_{01}}\left(\sum_{i=1}^{N}\sum_{q=1}^{Q}\sum_{h=1}^{H}\pi_{qhi}\frac{\partial \log f(y_{itk}|\Theta)}{\partial\gamma_1}\right) \\
&= \frac{\partial}{\partial\gamma_{01}}\left(\sum_{i=1}^{N}\sum_{q=1}^{Q}\sum_{h=1}^{H}\pi_{qhi}\frac{\partial \log f(y_{itk}|\Theta)}{\partial\gamma_1}\right) \\
&= \left(\sum_{i=1}^{N}\sum_{q=1}^{Q}\sum_{h=1}^{H}\pi_{qhi}\frac{\partial^2 \log f(y_{itk}|\Theta)}{\partial\gamma_{01}\partial\gamma_1} + \frac{\partial \log f(y_{itk}|\Theta)}{\partial\gamma_1}\frac{\partial\pi_{qhi}}{\partial\gamma_{01}}\right) \\
&= \sum_{i=1}^{N}\sum_{q=1}^{Q}\sum_{h=1}^{H}\left(\pi_{qhi}\frac{\partial^2 \log f(y_{itk}|\Theta)}{\partial\gamma_{01}\partial\gamma_1} + \frac{\partial \log f(y_{itk}|\Theta)}{\partial\gamma_1}\right. \\
&\quad \left.\left(\pi_{qhi}\frac{\partial \log f(y_{itk}|\Theta)}{\partial\gamma_{01}} - \pi_{qhi}\sum_{q=1}^{Q}\sum_{h=1}^{H}\pi_{rli}\frac{\partial \log f(y_{itk}|\Theta)}{\partial\gamma_{01}}\right)\right) \\
&= \sum_{i=1}^{N}\sum_{q=1}^{Q}\sum_{h=1}^{H}\pi_{qhi}\frac{\partial^2 \log f(y_{itk}|\Theta)}{\partial\gamma_{01}\partial\gamma_1} + \\
&\quad \pi_{qhi}\frac{\partial \log f(y_{itk}|\Theta)}{\partial\gamma_1}\frac{\partial \log f(y_{itk}|\Theta)}{\partial\gamma_{01}} \\
&\quad -\pi_{qhi}\frac{\partial \log f(y_{itk}|\Theta)}{\partial\gamma_1}\sum_{q=1}^{Q}\sum_{h=1}^{H}\pi_{rli}\frac{\partial \log f(y_{itk}|\Theta)}{\partial\gamma_{01}}
\end{aligned} \tag{3}
$$

The normal distribution with density $f(x)$ (mean μ and standard deviation $\sigma > 0$) has the following properties:

$\phi'(x) = -x\phi(x)$

$$\phi(A)\frac{1}{\sqrt{1-\rho^2}}\phi\left(\frac{B-\rho A}{\sqrt{1-\rho^2}}\right) = \phi(A,\ B,\ \rho)$$

$$\phi(A)\frac{1}{\sqrt{1-\rho^2}}\phi\left(\frac{B-\rho A}{\sqrt{1-\rho^2}}\right) = \phi(A,\ B,\ \rho)$$

$$\frac{\partial^2\Phi(A,B)}{\partial A\partial B} = \phi(A,B,\rho)$$

$$\frac{\partial^2\Phi(A,B)}{\partial\rho^2} = \frac{AB+\rho-\frac{\rho\left(A^2+B^2-2AB\rho\right)}{1-\rho^2}}{1-\rho^2}\phi(A,B,\rho)$$

$$\frac{\partial^2\Phi(A,B)}{\partial A\partial\rho} = \frac{\rho B-A}{1-\rho^2}\phi(A,B,\rho)$$

$$\frac{\partial^2\Phi(A,B)}{\partial r^2} = \frac{\partial^2\Phi(A,B)}{\partial\rho^2}\left(\frac{\partial\rho}{\partial r}\right)^2 + \frac{\partial\Phi(A,B)}{\partial\rho}\frac{\partial^2\rho}{\partial r^2} \text{ where } \frac{\partial^2\rho}{\partial r^2} = \frac{8\exp(2r)-8\exp(4r)}{(1+\exp(2r))^3}$$

$$\frac{\partial^2\Phi(A,B)}{\partial\alpha^2} = \frac{\partial^2\Phi(A,B)}{\partial\alpha^2}\left(\frac{\partial\alpha}{\partial\tau}\right)^2 + \frac{\partial\Phi(A,B)}{\partial\alpha}\frac{\partial^2\tau}{\partial\alpha^2} \text{ where } \frac{\partial^2\tau}{\partial\alpha^2} = \exp(\alpha)$$

$$\frac{\partial^2\Phi(A,B)}{\partial d^2} = \frac{\partial^2\Phi(A,B)}{\partial\sigma^2}(\sigma)^2 + \frac{\partial\Phi(A,B)}{\partial\sigma}\sigma$$

Let

$$\begin{aligned} f = \pi^{-1}&\sqrt{\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}exp\left\{2\rho_{\delta_{1i}\delta_{2i}}a_q a_h\right\} \\ &\prod_{t=1}^{T}\prod_{k=0}^{K}\{\left[\left(\left(1-\Phi\left(u_{1,qh}\right)\right)+\Phi\left(u_{1,qh},\ u_{1,qh0},-\rho_{ebeo}\right)\right)\right] \\ &\left[\Phi\left(u_{1,qh},\ u_{1,qhk},-\rho_{ebeo}\right)-\Phi\left(u_{1,qh},\ u_{1,qhk-1},-\rho_{ebeo}\right)\right] \\ &\left[\Phi\left(u_{1,qh},\ u_{1,qK}-\tau_{K-1},\rho_{ebeo}\right)\right]\}^{d_{itk}}\Big\} \end{aligned} \tag{4}$$

The second differentiation of (4.58) with respect to θ is given by,

$$\frac{\partial^2 \log f}{\partial \theta' \partial \theta} = \sum_{t=1}^{T} \sum_{k=0}^{K} d_{itk} \frac{1}{C} \left\{ \left[u_{1,qh} \varphi \left(u_{1,qh} \right) + \right. \right.$$
$$\left. + \left(-u_{1,qh} \varphi \left(u_{1,qh} \right) \Phi \left(\frac{u_{1,qh0} + \rho_{ebeo} u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}} \right) + \rho_{ebeo} \varphi \left(u_{1,qh0},\ u_{1,qh},\ -\rho_{ebeo} \right) \right) \right]$$
$$\left[\left(-u_{1,qh} \varphi \left(u_{1,qh} \right) \Phi \left(\frac{u_{1,qhk} + \rho_{ebeo} u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}} \right) + \rho_{ebeo} \varphi \left(u_{1,qhk},\ u_{1,qh},\ -\rho_{ebeo} \right) \right) - \right.$$
$$\left. \left(-u_{1,qh} \varphi \left(u_{1,qh} \right) \Phi \left(\frac{u_{1,qhk-1} + \rho_{ebeo} u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}} \right) + \rho_{ebeo} \varphi \left(u_{1,qhk-1},\ u_{1,qh},\ -\rho_{ebeo} \right) \right) \right]$$
$$\left. + \left[-u_{1,qh} \varphi \left(u_{1,qh} \right) \Phi \left(\frac{u_{1,qhK} - \rho_{ebeo} u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}} \right) - \rho_{ebeo} \varphi \left(u_{1,qhK},\ u_{1,qh},\ \rho_{ebeo} \right) \right] \right\} XX'$$
$$- \frac{\partial \log f}{\partial \theta} \frac{\partial \log f}{\partial \theta} \tag{5}$$

The second differentiation of (4.58) with respect to d_1 is given by,

$$\frac{\partial^2 \log f}{\partial d_1 \partial \theta} = \sum_{t=1}^{T} \sum_{k=0}^{K} d_{itk} \frac{1}{C} \left\{ u_{1,qh} \varphi \left(u_{1,qh} \right) \right.$$
$$\left. + \left(-u_{1,qh} \varphi \left(u_{1,qh} \right) \Phi \left(\frac{u_{1,qh0} + \rho_{ebeo} u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}} \right) + \rho_{ebeo} \varphi \left(u_{1,qh0},\ u_{1,qh},\ -\rho_{ebeo} \right) \right) \right]$$
$$\left[\left(-u_{1,qh} \varphi \left(u_{1,qh} \right) \Phi \left(\frac{u_{1,qhk} + \rho_{ebeo} u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}} \right) + \rho_{ebeo} \varphi \left(u_{1,qhk},\ u_{1,qh},\ -\rho_{ebeo} \right) \right) - \right.$$
$$\left. \left(-u_{1,qh} \varphi \left(u_{1,qh} \right) \Phi \left(\frac{u_{1,qhk-1} + \rho_{ebeo} u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}} \right) + \rho_{ebeo} \varphi \left(u_{1,qhk-1},\ u_{1,qh},\ -\rho_{ebeo} \right) \right) \right]$$
$$\left. + \left[-u_{1,qh} \varphi \left(u_{1,qh} \right) \Phi \left(\frac{u_{1,qhK} - \rho_{ebeo} u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}} \right) - \rho_{ebeo} \varphi \left(u_{1,qhK},\ u_{1,qh},\ \rho_{ebeo} \right) \right] \right\}$$
$$X a_h e^{\log(\sigma_1)} \sqrt{2 \left(1 - \rho_{\delta_{1i} \delta_{2i}}^2 \right)} - \frac{\partial \log f}{\partial d_1} \frac{\partial \log f}{\partial \theta} \tag{6}$$

The second differentiation of (4.58) with respect to ψ is given by,

$$\frac{\partial^2 \log f}{\partial \psi \partial \theta} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[\varphi\left(u_{1,qh}, u_{1,qh0}, -\rho_{ebeo}\right)\right] + \right.$$
$$\left[\varphi\left(u_{1,qh}, u_{1,qhk}, -\rho_{ebeo}\right) - \varphi\left(u_{1,qh}, u_{1,qhk-1}, -\rho_{ebeo}\right)\right]$$
$$\left. + \left[-\varphi\left(u_{1,qh}, u_{1,qhK}, \rho_{ebeo}\right)\right]\right\} XZ - \frac{\partial \log f}{\partial \psi}\frac{\partial \log f}{\partial \theta}$$
$$XZ - \frac{\partial \log f}{\partial \psi}\frac{\partial \log f}{\partial \theta} \tag{7}$$

The second differentiation of (4.58) with respect to d_2 is given by,

$$\frac{\partial^2 \log f}{\partial d_2 \partial \theta} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[\varphi\left(u_{1,qh}, u_{1,qh0}, -\rho_{ebeo}\right)\right] + \right.$$
$$\left[\varphi\left(u_{1,qh}, u_{1,qhk}, -\rho_{ebeo}\right) - \varphi\left(u_{1,qh}, u_{1,qhk-1}, -\rho_{ebeo}\right)\right] +$$
$$\left.\left[-\varphi\left(u_{1,qh}, u_{1,qhK}, \rho_{ebeo}\right)\right]\right\} X\left(-a_q \mathrm{e}^{\log(\sigma_2)}\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)}\right)$$
$$-\frac{\partial \log f}{\partial d_2}\frac{\partial \log f}{\partial \theta} \tag{8}$$

The second differentiation of (4.58) with respect to r_e is given by,

$$\frac{\partial^2 \log f\left(y_{itk}|\Theta\right)}{\partial r_e \partial \theta} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[\frac{\rho_{ebeo}u_{1,qh0} + u_{1,qh}}{1-\rho^2_{ebeo}}\varphi\left(u_{1,qh0}, u_{1,qh}, -\rho_{ebeo}\right)\right] + \right.$$
$$\left[\frac{\rho_{ebeo}u_{1,qhk} + u_{1,qh}}{1-\rho^2_{ebeo}}\varphi\left(u_{1,qhk}, u_{1,qh}, -\rho_{ebeo}\right) - \right.$$
$$\left.\frac{\rho_{ebeo}u_{1,qhk-1} + u_{1,qh}}{1-\rho^2_{ebeo}}\varphi\left(u_{1,qhk-1}, u_{1,qh}, -\rho_{ebeo}\right)\right]$$
$$\left. + \left[\frac{\rho_{ebeo}u_{1,qhK} - u_{1,qh}}{1-\rho^2_{ebeo}}\varphi\left(u_{1,qhK}, u_{1,qh}, \rho_{ebeo}\right)\right]\right\} X\frac{4\exp(2r_e)}{(1+\exp(2r_e))^2}$$
$$-\frac{\partial \log f}{\partial r_e}\frac{\partial \log f}{\partial \theta}$$
$$\tag{9}$$

The second differentiation of (4.58) with respect to r is given by,

$$\frac{\partial^2 \log f}{\partial r \partial \theta} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[-u_{1,qh}\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b} + \left(u_{1,qh}\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}\Phi\left(\frac{u_{1,qh0}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right) + \left(-a_h\sigma_{\delta_1}\rho_{ebeo}+a_q\sigma_{\delta_2}\right)\varphi\left(u_{1,qh0},u_{1,qh},-\rho_{ebeo}\right)\right] \left[\left(u_{1,qh}\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}\Phi\left(\frac{u_{1,qhk}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right) + \left(-a_h\sigma_{\delta_1}\rho_{ebeo}+a_q\sigma_{\delta_2}\right)\varphi\left(u_{1,qhk},u_{1,qh},-\rho_{ebeo}\right)\right) \left(u_{1,qh}\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}\Phi\left(\frac{u_{1,qhk-1}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right) + \left(-a_h\sigma_{\delta_1}\rho_{ebeo}+a_q\sigma_{\delta_2}\right)\varphi\left(u_{1,qhk-1},u_{1,qh},-\rho_{ebeo}\right)\right)\right] \left[u_{1,qh}\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}\Phi\left(\frac{u_{1,qhK}-\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right) + \left(-a_q\sigma_{\delta_o}+\rho_{ebeo}a_h\sigma_{\delta_b}\right)\varphi\left(u_{1,qh},u_{1,qhK},\rho_{ebeo}\right)\right] X\frac{\sqrt{2}\rho_{\delta_{1i}\delta_{2i}}}{\sqrt{\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}}\frac{4\exp(2r)}{\left(1+\exp(2r)\right)^2} - \frac{\partial \log f}{\partial r}\frac{\partial \log f}{\partial \theta} \tag{10}$$

The second differentiation of (4.58) with respect to α_1 is given by,

$$\frac{\partial^2 \log f}{\partial \alpha_1 \partial \theta} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[\varphi\left(u_{1,qh1},\ u_{1,qh},\ -\rho_{ebeo}\right)\right] + \left[-\varphi\left(u_{1,qh1},\ u_{1,qh},\ -\rho_{ebeo}\right)\right]\right\}X - \frac{\partial \log f}{\partial \alpha_1}\frac{\partial \log f}{\partial \theta} \tag{11}$$

The second derivative of equation (4.58) with respect to α_k is given by,

$$\frac{\partial^2 \log f}{\partial \alpha_k \partial \theta} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[\varphi\left(u_{1,qhk},\ u_{1,qh},\ -\rho_{ebeo}\right)\right] + \left[-\varphi\left(u_{1,qhk},\ u_{1,qh},\ -\rho_{ebeo}\right)\right]\right\}e^{\alpha_k}X - \frac{\partial \log f}{\partial \alpha_k}\frac{\partial \log f}{\partial \theta} \tag{12}$$

The second differentiation of (4.60) with respect to d_1 is given by,

$$\frac{\partial^2 \log f}{\partial d_1^2} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[u_{1,qh}\varphi\left(u_{1,qh}\right)+\right.\right.$$
$$\left.\left(-u_{1,qh}\varphi\left(u_{1,qh}\right)\Phi\left(\frac{u_{1,qh0}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)+\rho_{ebeo}\varphi\left(u_{1,qh0},\ u_{1,qh},\ -\rho_{ebeo}\right)\right)\right]$$
$$+\left[\left(-u_{1,qh}\varphi\left(u_{1,qh}\right)\Phi\left(\frac{u_{1,qhk}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)+\rho_{ebeo}\varphi\left(u_{1,qhk},\ u_{1,qh},\ -\rho_{ebeo}\right)\right)-\right.$$
$$\left.\left(-u_{1,qh}\varphi\left(u_{1,qh}\right)\Phi\left(\frac{u_{1,qhk-1}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)+\rho_{ebeo}\varphi\left(u_{1,qhk-1},\ u_{1,qh},\ -\rho_{ebeo}\right)\right)\right]$$
$$\left.+\left[-u_{1,qh}\varphi\left(u_{1,qh}\right)\Phi\left(\frac{u_{1,qhK}-\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)-\rho_{ebeo}\varphi\left(u_{1,qhK},\ u_{1,qh},\ \rho_{ebeo}\right)\right]\right\}$$
$$\left(a_h\sqrt{2\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}e^{\log(\sigma_1)}\right)^2+\left\{\left[-\varphi\left(u_{1,qh}\right)+\varphi\left(u_{1,qh}\right)\Phi\left(\frac{u_{1,qh0}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right]\right.$$
$$+\left[\varphi\left(u_{1,qh}\right)\Phi\left(\frac{u_{1,qhk}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)-\varphi\left(u_{1,qh}\right)\Phi\left(\frac{u_{1,qhk-1}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right]$$
$$\left.+\left[\varphi\left(u_{1,qh}\right)\Phi\left(\frac{u_{1,qhK}-\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right]\right\}a_h e^{\log(\sigma_1)}\sqrt{2\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}$$
$$-\frac{\partial \log f}{\partial d_1}\frac{\partial \log f}{\partial d_1} \tag{13}$$

The second differentiation of (4.60) with respect to ψ is given by,

$$\frac{\partial^2 \log f}{\partial \psi \partial d_1} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[\varphi\left(u_{1,qh},u_{1,qh0},-\rho_{ebeo}\right)\right]+\right.$$
$$\left[\varphi\left(u_{1,qh},u_{1,qhk},-\rho_{ebeo}\right)-\varphi\left(u_{1,qh},u_{1,qhk-1},-\rho_{ebeo}\right)\right]+$$
$$\left.\left[-\varphi\left(u_{1,qh},u_{1,qhK},\rho_{ebeo}\right)\right]\right\}a_h e^{\log(\sigma_1)}\sqrt{2\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}Z$$
$$-\frac{\partial \log f}{\partial \psi_1}\frac{\partial \log f}{\partial d_1} \tag{14}$$

The second differentiation of (4.60) with respect to d_2 is given by,

$$\frac{\partial^2 \log f}{\partial d_2 \partial d_1} = \sum_{t=1}^{T} \sum_{k=0}^{K} d_{itk} \frac{1}{C} \left\{ \left[\varphi \left(u_{1,qh}, u_{1,qh0}, -\rho_{ebeo} \right) \right] + \right.$$
$$\left[\varphi \left(u_{1,qh}, u_{1,qhk}, -\rho_{ebeo} \right) - \varphi \left(u_{1,qh}, u_{1,qhk-1}, -\rho_{ebeo} \right) \right]$$
$$\left. -\varphi \left(u_{1,qh}, u_{1,qhK}, \rho_{ebeo} \right) \right\} a_h e^{\log(\sigma_1)} \sqrt{2 \left(1 - \rho^2_{\delta_{1i}\delta_{2i}} \right)}$$
$$\left(-a_q e^{\log(\sigma_2)} \sqrt{2 \left(1 - \rho^2_{\delta_{1i}\delta_{2i}} \right)} \right) - \frac{\partial \log f}{\partial d_2} \frac{\partial \log f}{\partial d_1} \tag{15}$$

The second differentiation of (4.60) with respect to r_e is given by,

$$\frac{\partial^2 \log f}{\partial r_e \partial d_1} = \sum_{t=1}^{T} \sum_{k=0}^{K} d_{itk} \frac{1}{C} \left\{ \left[\frac{u_{1,qh} + \rho_{ebeo} u_{1,qh0}}{1 - \rho^2_{ebeo}} \varphi \left(u_{1,qh0}, u_{1,qh}, -\rho_{ebeo} \right) \right] \right.$$
$$\left[\frac{u_{1,qh} + \rho_{ebeo} u_{1,qhk}}{1 - \rho^2_{ebeo}} \varphi \left(u_{1,qhk}, u_{1,qh}, -\rho_{ebeo} \right) - \right.$$
$$\left. - \frac{u_{1,qh} + \rho_{ebeo} u_{1,qhk-1}}{1 - \rho^2_{ebeo}} \varphi \left(u_{1,qhk-1}, u_{1,qh}, -\rho_{ebeo} \right) \right]$$
$$\left. + \frac{\rho_{ebeo} u_{1,qhK} - u_{1,qh}}{1 - \rho^2_{ebeo}} \varphi \left(u_{1,qhK}, u_{1,qh}, \rho_{ebeo} \right) \right\} a_h e^{\log(\sigma_1)} \sqrt{2 \left(1 - \rho^2_{\delta_{1i}\delta_{2i}} \right)}$$
$$\frac{4 \exp(2 r_e)}{(1 + \exp(2 r_e))^2} - \frac{\partial \log f}{\partial r_e} \frac{\partial \log f}{\partial d_1} \tag{16}$$

The second differentiation of (4.60) with respect to r is given by,

$$
\begin{aligned}
\frac{\partial^2 \log f}{\partial r \partial d_1} = & \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[-u_{1,qh}\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}\right.\right. \\
& + \left(u_{1,qh}\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}\Phi\left(\frac{u_{1,qh0}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right) \\
& \left.+\left(a_q\sigma_{\delta_o}-\rho_{ebeo}a_h\sigma_{\delta_b}\right)\varphi\left(u_{1,qh0},u_{1,qh},-\rho_{ebeo}\right)\right]+ \\
& \left[\left(u_{1,qh}\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}\Phi\left(\frac{u_{1,qhk}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right.\right. \\
& \left.+\left(a_q\sigma_{\delta_o}-a_h\sigma_{\delta_b}\rho_{ebeo}\right)\varphi\left(u_{1,qhk},u_{1,qh},-\rho_{ebeo}\right)\right) \\
& -\left(u_{1,qh}\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}\Phi\left(\frac{u_{1,qhk-1}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right. \\
& \left.\left.+\left(a_q\sigma_{\delta_o}-a_h\sigma_{\delta_b}\rho_{ebeo}\right)\varphi\left(u_{1,qhk-1},u_{1,qh},-\rho_{ebeo}\right)\right)\right] \\
& -\left[u_{1,qh}\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}\Phi\left(\frac{u_{1,qhK}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)+\right. \\
& \left.\left.\left(-a_q\sigma_{\delta_o}+a_h\sigma_{\delta_b}\rho_{ebeo}\right)\varphi\left(u_{1,qh},u_{1,qhK},\rho_{ebeo}\right)\right]\right\}a_h e^{\log(\sigma_1)} \\
& \sqrt{2\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}\frac{\sqrt{2}\rho_{\delta_{1i}\delta_{2i}}}{\sqrt{\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}}\frac{4\exp(2r)}{\left(1+\exp(2r)\right)^2}+ \\
& \left\{\left[-\varphi\left(u_{1,qh}\right)+\varphi\left(u_{1,qh}\right)\Phi\left(\frac{u_{1,qh0}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right]+\right. \\
& \left[\varphi\left(u_{1,qh}\right)\Phi\left(\frac{u_{1,qhk}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)-\varphi\left(u_{1,qh}\right)\Phi\left(\frac{u_{1,qhk-1}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right] \\
& \left.+\left[\varphi\left(u_{1,qh}\right)\Phi\left(\frac{u_{1,qhK}-\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right]\right\}a_h e^{\log(\sigma_1)}\frac{-\sqrt{2}\rho_{\delta_{1i}\delta_{2i}}}{\sqrt{\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}} \\
& \frac{4\exp(2r)}{\left(1+\exp(2r)\right)^2}-\frac{\partial \log f}{\partial r_1}\frac{\partial \log f}{\partial d_1}
\end{aligned}
\tag{17}
$$

The second differentiation of (4.60) with respect to α_1 is given by,

$$\frac{\partial^2 \log f}{\partial \alpha_1 \partial d_1} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[\varphi\left(u_{1,qh1},\ u_{1,qh},\ -\rho_{ebeo}\right)\right] + \left[-\varphi\left(u_{1,qh1},\ u_{1,qh},\ -\rho_{ebeo}\right)\right]\right\} a_h e^{\log(\sigma_1)}\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)} \frac{\partial \log f}{\partial \alpha_{11}}\frac{\partial \log f}{{}_1\partial d_1} \tag{18}$$

The second differentiation of (4.60) with respect to α_k is given by,

$$\frac{\partial^2 \log f}{\partial \alpha_k \partial d_1} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[\varphi\left(u_{1,qhk},\ u_{1,qh},\ -\rho_{ebeo}\right)\right] + \left[-\varphi\left(u_{1,qhk},\ u_{1,qh},\ -\rho_{ebeo}\right)\right]\right\} \mathrm{e}^{\alpha_k} a_h e^{\log(\sigma_1)}\sqrt{2\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)} - \frac{\partial \log f}{\partial \alpha_k}\frac{\partial \log f}{\partial d_1} \tag{19}$$

The second differentiation of (4.61) with respect to ψ is given by,

$$\begin{aligned}\frac{\partial^2 \log f}{\partial \psi' \partial \psi} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\Bigg\{ & \\ &\left[-u_{1,qh0}\varphi\left(u_{1,qh0}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qh0}}{\sqrt{1-\rho^2_{ebeo}}}\right) + \rho_{ebeo}\varphi\left(u_{1,qh0},\ u_{1,qh},\ -\rho_{ebeo}\right)\right] + \\ &\Bigg[\left(-u_{1,qhk}\varphi\left(u_{1,qhk}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk}}{\sqrt{1-\rho^2_{ebeo}}}\right) + \rho_{ebeo}\varphi\left(u_{1,qhk},\ u_{1,qh},\ -\rho_{ebeo}\right)\right) \\ &- \left(-u_{1,qhk-1}\varphi\left(u_{1,qhk-1}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk-1}}{\sqrt{1-\rho^2_{ebeo}}}\right)\right. \\ &\left. + \rho_{ebeo}\varphi\left(u_{1,qhk-1},\ u_{1,qh},\ -\rho_{ebeo}\right)\right)\Bigg] + \\ &\left[-u_{1,qhK}\varphi\left(u_{1,qhK}\right)\Phi\left(\frac{u_{1,qh}-\rho_{ebeo}u_{1,qhK}}{\sqrt{1-\rho^2_{ebeo}}}\right) - \rho_{ebeo}\varphi\left(u_{1,qhK},\ u_{1,qh},\ \rho_{ebeo}\right)\right]\Bigg\} \\ & ZZ' - \frac{\partial \log f}{\partial \psi}\frac{\partial \log f}{\partial \psi}\end{aligned} \tag{20}$$

The second differentiation of (4.61) with respect to d_2 is given by,

$$\frac{\partial^2 \log f}{\partial d_2 \partial \psi} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\{$$

$$\left[-u_{1,qh0}\varphi\left(u_{1,qh0}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qh0}}{\sqrt{1-\rho_{ebeo}^2}}\right)+\rho_{ebeo}\varphi\left(u_{1,qh0},\ u_{1,qh},\ -\rho_{ebeo}\right)\right]+$$

$$\left[\left(-u_{1,qhk}\varphi\left(u_{1,qhk}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk}}{\sqrt{1-\rho_{ebeo}^2}}\right)+\rho_{ebeo}\varphi\left(u_{1,qhk},\ u_{1,qh},\ -\rho_{ebeo}\right)\right)\right.$$

$$-\left(-u_{1,qhk-1}\varphi\left(u_{1,qhk-1}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk-1}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right.$$

$$\left.\left.+\rho_{ebeo}\varphi\left(u_{1,qhk-1},\ u_{1,qh},\ -\rho_{ebeo}\right)\right)\right]+$$

$$\left.\left[-u_{1,qhK}\varphi\left(u_{1,qhK}\right)\Phi\left(\frac{u_{1,qh}-\rho_{ebeo}u_{1,qhK}}{\sqrt{1-\rho_{ebeo}^2}}\right)-\rho_{ebeo}\varphi\left(u_{1,qhK},\ u_{1,qh},\ \rho_{ebeo}\right)\right]\right\}$$

$$Z\left(-a_q e^{\log(\sigma_2)}\sqrt{2\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}\right)-\frac{\partial \log f}{\partial d_2}\frac{\partial \log f}{\partial \psi} \tag{21}$$

The second differentiation of (4.61) with respect to r_e is given by,

$$\frac{\partial^2 \log f}{\partial r_e \partial \psi} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[\frac{u_{1,qh0}+\rho_{ebeo}u_{1,qh}}{1-\rho_{ebeo}^2}\varphi\left(u_{1,qh0},u_{1,qh},-\rho_{ebeo}\right)\right]+\right.$$

$$\left[\left(\frac{u_{1,qhk}+\rho_{ebeo}u_{1,qh}}{1-\rho_{ebeo}^2}\varphi\left(u_{1,qhk},u_{1,qh},-\rho_{ebeo}\right)\right)-\right.$$

$$\left.-\frac{u_{1,qhk-1}+\rho_{ebeo}u_{1,qh}}{1-\rho_{ebeo}^2}\varphi\left(u_{1,qhk-1},u_{1,qh},-\rho_{ebeo}\right)\right]+$$

$$\left.\left[-\frac{u_{1,qh}\rho_{ebeo}-u_{1,qhK}}{1-\rho_{ebeo}^2}\varphi\left(u_{1,qhK},u_{1,qh},\rho_{ebeo}\right)\right]\right\}Z\frac{4\exp(2r_e)}{(1+\exp(2r_e))^2}$$

$$-\frac{\partial \log f}{\partial r_e}\frac{\partial \log f}{\partial \psi} \tag{22}$$

The second differentiation of (4.61) with respect to r is given by,

$$\frac{\partial^2 \log f}{\partial r \partial \psi} = \sum_{t=1}^{T} \sum_{k=0}^{K} d_{itk} \frac{1}{C} \{$$
$$\left[\left(-u_{1,qh0} \varphi\left(u_{1,qh0}\right) a_q \sigma_{\delta_o} \Phi \left(\frac{u_{1,qh} + \rho_{ebeo} u_{1,qh0}}{\sqrt{1-\rho_{ebeo}^2}} \right) \right) \right.$$
$$\left. + \left(-a_h \sigma_{\delta_1} + a_q \sigma_{\delta_o} \rho_{ebeo}\right) \varphi\left(u_{1,qh0}, u_{1,qh}, -\rho_{ebeo}\right) \right]$$
$$\left[\left(-u_{1,qhk} \varphi\left(u_{1,qhk}\right) a_q \sigma_{\delta_o} \Phi \left(\frac{u_{1,qh} + \rho_{ebeo} u_{1,qhk}}{\sqrt{1-\rho_{ebeo}^2}} \right) \right. \right.$$
$$\left. + \left(-a_h \sigma_{\delta_1} + a_q \sigma_{\delta_o} \rho_{ebeo}\right) \varphi\left(u_{1,qhk}, u_{1,qh}, -\rho_{ebeo}\right) \right) -$$
$$\left(-u_{1,qhk-1} \varphi\left(u_{1,qhk-1}\right) a_q \sigma_{\delta_o} \Phi \left(\frac{u_{1,qh} + \rho_{ebeo} u_{1,qhk-1}}{\sqrt{1-\rho_{ebeo}^2}} \right) \right.$$
$$\left. \left. + \left(-a_h \sigma_{\delta_1} + a_q \sigma_{\delta_o} \rho_{ebeo}\right) \varphi\left(u_{1,qhk-1}, u_{1,qh}, -\rho_{ebeo}\right) \right) \right]$$
$$+ \left[-u_{1,qhK} \varphi\left(u_{1,qhK}\right) a_q \sigma_{\delta_o} \Phi \left(\frac{u_{1,qh} - \rho_{ebeo} u_{1,qhK}}{\sqrt{1-\rho_{ebeo}^2}} \right) \right.$$
$$\left. - \left(-a_h \sigma_{\delta_b} + \rho_{ebeo} a_q \sigma_{\delta_o}\right) \varphi\left(u_{1,qh}, u_{1,qhK}, \rho_{ebeo}\right) \right] \}$$
$$Z \frac{\sqrt{2} \rho_{\delta_{1i}\delta_{2i}}}{\sqrt{\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}} \frac{4\exp(2r)}{(1+\exp(2r))^2} - \frac{\partial \log f}{\partial r} \frac{\partial \log f}{\partial \psi} \tag{23}$$

The second differentiation of (4.61) with respect to α_1 is given by,

$$\frac{\partial^2 \log f}{\partial \alpha_1 \partial \psi} = \sum_{t=1}^{T} \sum_{k=0}^{K} d_{itk} \frac{1}{C}$$
$$\left\{ \left[-u_{1,qh1} \varphi\left(u_{1,qh1}\right) \Phi \left(\frac{u_{1,qh} + \rho_{ebeo} u_{1,qh1}}{\sqrt{1-\rho_{ebeo}^2}} \right) + \rho_{ebeo} \varphi\left(u_{1,qh1},\ u_{1,qh},\ -\rho_{ebeo}\right) \right] - \right.$$
$$\left. \left[\left(-u_{1,qh1} \varphi\left(u_{1,qh1}\right) \Phi \left(\frac{u_{1,qh} + \rho_{ebeo} u_{1,qh1}}{\sqrt{1-\rho_{ebeo}^2}} \right) + \rho_{ebeo} \varphi\left(u_{1,qh1},\ u_{1,qh},\ -\rho_{ebeo}\right) \right) \right] \right\}$$
$$Z - \frac{\partial \log f}{\partial \alpha_1} \frac{\partial \log f}{\partial \psi} \tag{24}$$

The second differentiation of (4.61) with respect to α_k is given by,

$$\frac{\partial^2 \log f}{\partial \alpha_k \partial \psi} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}$$

$$\left\{\left[-u_{1,qhk}\varphi\left(u_{1,qhk}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk}}{\sqrt{1-\rho_{ebeo}^2}}\right)+\rho_{ebeo}\varphi\left(u_{1,qhk},\ u_{1,qh},\ -\rho_{ebeo}\right)\right]+\right.$$

$$\left.\left[-\left(-u_{1,qhk}\varphi\left(u_{1,qhk}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk}}{\sqrt{1-\rho_{ebeo}^2}}\right)+\rho_{ebeo}\varphi\left(u_{1,qhk},\ u_{1,qh},\ -\rho_{ebeo}\right)\right)\right]\right\}$$

$$Z-\frac{\partial \log f}{\partial \alpha_k}\frac{\partial \log f}{\partial \psi} \tag{25}$$

The second differentiation of (4.62) with respect to d_2 is given by,

$$\frac{\partial^2 \log f}{\partial d'_2 \partial d_2} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\{$$

$$\left[-u_{1,qh0}\varphi\left(u_{1,qh0}\right)\Phi\left(\frac{u_{1,qh0}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)+\rho_{ebeo}\varphi\left(u_{1,qh0},\ u_{1,qh},\ -\rho_{ebeo}\right)\right]+$$

$$\left[\left(-u_{1,qhk}\varphi\left(u_{1,qh0}\right)\Phi\left(\frac{u_{1,qhk}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)+\rho_{ebeo}\varphi\left(u_{1,qhk},\ u_{1,qh},\ -\rho_{ebeo}\right)\right)-\right.$$

$$\left.\left(-u_{1,qhk-1}\varphi\left(u_{1,qh0}\right)\Phi\left(\frac{u_{1,qhk-1}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)+\rho_{ebeo}\varphi\left(u_{1,qhk-1},\ u_{1,qh},\ -\rho_{ebeo}\right)\right)\right]$$

$$\left.+\left[-u_{1,qhK}\varphi\left(u_{1,qh0}\right)\Phi\left(\frac{u_{1,qhk-1}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)-\rho_{ebeo}\varphi\left(u_{1,qhK},\ u_{1,qh},\ \rho_{ebeo}\right)\right]\right\}$$

$$\left(-a_q\sqrt{2\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}e^{\log(\sigma_2)}\right)^2+\left\{\left[\varphi\left(u_{1,qh0}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qh0}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right]+\right.$$

$$\left[\left(\varphi\left(u_{1,qhk}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right)-\left(\varphi\left(u_{1,qhk-1}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk-1}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right)\right]$$

$$\left.\left[-\varphi\left(u_{1,qhK}\right)\Phi\left(\frac{u_{1,qh}-\rho_{ebeo}u_{1,qhK}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right]\right\}\left(-a_q e^{\log(\sigma_2)}\sqrt{2\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}\right)$$

$$-\frac{\partial \log f}{\partial \partial d_2}\frac{\partial \log f}{\partial d_2} \tag{26}$$

The second differentiation of (4.62) with respect to r_e is given by,

$$\frac{\partial^2 \log f}{\partial r_e \partial d_2} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\Bigg\{\left[\frac{u_{1,qh}+\rho_{ebeo}u_{1,qh0}}{1-\rho_{ebeo}^2}\varphi\left(u_{1,qh0},u_{1,qh},-\rho_{ebeo}\right)\right]+$$
$$\left[\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk}}{1-\rho_{ebeo}^2}\varphi\left(u_{1,qhk},u_{1,qh},-\rho_{ebeo}\right)\right)-\right.$$
$$\left.\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk-1}}{1-\rho_{ebeo}^2}\varphi\left(u_{1,qhk-1},u_{1,qh},-\rho_{ebeo}\right)\right]$$
$$-\left[\left[\frac{\rho_{ebeo}u_{1,qh0}-u_{1,qh}+}{1-\rho_{ebeo}^2}\varphi\left(u_{1,qhK},u_{1,qh},\rho_{ebeo}\right)\right]\right]\Bigg\}$$
$$\left(-a_q e^{\log(\sigma_2)}\sqrt{2\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}\right)\frac{4\exp(2r_e)}{(1+\exp(2r_e))^2}-\frac{\partial \log f}{\partial r_e}\frac{\partial \log f}{\partial d_2} \tag{27}$$

The second differentiation of (4.62) with respect to r is given by,

$$\frac{\partial^2 \log f}{\partial r \partial d_2} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\{$$
$$\left[\left(-u_{1,qh0}\varphi\left(u_{1,qh0}\right)a_h\sigma_{\delta_b}\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qh0}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right)\right.$$
$$\left.+\left(-a_h\sigma_{\delta_1}+a_q\sigma_{\delta_2}\rho_{ebeo}\right)\varphi\left(u_{1,qh0},u_{1,qh},-\rho_{ebeo}\right)\right]$$
$$\left[\left(-u_{1,qhk}\varphi\left(u_{1,qhk}\right)a_h\sigma_{\delta_b}\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right.\right.$$
$$\left.+\left(-a_h\sigma_{\delta_1}+a_q\sigma_{\delta_2}\rho_{ebeo}\right)\varphi\left(u_{1,qhk},u_{1,qh},-\rho_{ebeo}\right)\right)$$
$$-\left(-u_{1,qhk-1}\varphi\left(u_{1,qhk-1}\right)a_h\sigma_{\delta_b}\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk-1}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right.$$
$$\left.\left.+\left(-a_h\sigma_{\delta_b}+\rho_{ebeo}a_q\sigma_{\delta_2}\right)\varphi\left(u_{1,qhk-1},u_{1,qh},-\rho_{ebeo}\right)\right)\right]$$
$$+\left[-u_{1,qhK}\varphi\left(u_{1,qhK}\right)a_h\sigma_{\delta_b}\Phi\left(\frac{u_{1,qh}-\rho_{ebeo}u_{1,qhK}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right.$$
$$\left.-\left(a_q\sigma_{\delta_o}\rho_{ebeo}-a_h\sigma_{\delta_b}\right)\varphi\left(u_{1,qh},u_{1,qhK},\rho_{ebeo}\right)\right]\}$$
$$\left(-a_q e^{\log(\sigma_2)}\sqrt{2\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}\right)\frac{\sqrt{2}\rho_{\delta_{1i}\delta_{2i}}}{\sqrt{\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}}\frac{4\exp(2r)}{(1+\exp(2r))^2}+$$
$$\left\{\left[\varphi\left(u_{1,qh0}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qh0}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right]+\right.$$
$$\left[\left(\varphi\left(u_{1,qhk}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right)-\left(\varphi\left(u_{1,qhk-1}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk-1}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right)\right]$$
$$\left.\left[-\varphi\left(u_{1,qhK}\right)\Phi\left(\frac{u_{1,qh}-\rho_{ebeo}u_{1,qhK}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right]\right\}\left(-a_q e^{\log(\sigma_2)}\right)\frac{-\sqrt{2}\rho_{\delta_{1i}\delta_{2i}}}{\sqrt{\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}}\frac{4\exp(2r)}{(1+\exp(2r))^2}$$
$$-\frac{\partial \log f}{\partial r}\frac{\partial \log f}{\partial d_2} \qquad (28)$$

The second differentiation of (4.62) with respect to α_1 is given by,

$$\frac{\partial^2 \log f}{\partial \alpha_1 \partial d_2} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\{$$

$$\left[-u_{1,qh1}\varphi\left(u_{1,qh1}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qh1}}{\sqrt{1-\rho_{ebeo}^2}}\right)+\rho_{ebeo}\varphi\left(u_{1,qh1},\ u_{1,qh},\ -\rho_{ebeo}\right)\right]+$$

$$\left[-\left(-u_{1,qh1}\varphi\left(u_{1,qh1}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qh1}}{\sqrt{1-\rho_{ebeo}^2}}\right)+\rho_{ebeo}\varphi\left(u_{1,qh1},\ u_{1,qh},\ -\rho_{ebeo}\right)\right)\right]\}$$

$$\left(-a_q e^{\log(\sigma_2)}\sqrt{2\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}\right)-\frac{\partial \log f}{\partial \alpha_{12}}\frac{\partial \log f}{\partial d_2} \tag{29}$$

The second differentiation of (4.62) with respect to α_k is given by,

$$\frac{\partial^2 \log f}{\partial \alpha_k \partial d_2} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\{$$

$$\left[-u_{1,qhk}\varphi\left(u_{1,qhk}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk}}{\sqrt{1-\rho_{ebeo}^2}}\right)+\rho_{ebeo}\varphi\left(u_{1,qhk},\ u_{1,qh},\ -\rho_{ebeo}\right)\right]+$$

$$\left[-\left(-u_{1,qhk}\varphi\left(u_{1,qhk}\right)\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk}}{\sqrt{1-\rho_{ebeo}^2}}\right)+\rho_{ebeo}\varphi\left(u_{1,qhk},\ u_{1,qh},\ -\rho_{ebeo}\right)\right)\right]\}$$

$$\left(-a_q e^{\log(\sigma_2)}\sqrt{2\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}\right)-\frac{\partial \log f}{\partial \alpha_k}\frac{\partial \log f}{\partial d_2} \tag{30}$$

The second differentiation of (4.65) with respect to r_e is given by,

$$\frac{\partial^2 \log f}{\partial r_e{}' \partial r_e} = \sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\{$$

$$\left[\left(\frac{-u_{1,qh}u_{1,qh0}+\rho_{ebeo}-\rho_{ebeo}\frac{u^2{}_{1,qh}+u^2{}_{1,qh0}+2\rho_{ebeo}u_{1,qh}u_{1,qh0}}{1-\rho^2{}_{ebeo}}}{1-\rho^2{}_{ebeo}}\right)\varphi\left(u_{1,qh},\ u_{1,qh0},-\rho_{ebeo}\right)\right]+$$

$$\left[\left(\frac{-u_{1,qh}u_{1,qhk}+\rho_{ebeo}-\rho_{ebeo}\frac{u^2{}_{1,qh}+u^2{}_{1,qhk}+2\rho_{ebeo}u_{1,qh}u_{1,qhk}}{1-\rho^2{}_{ebeo}}}{1-\rho^2{}_{ebeo}}\varphi\left(u_{1,qh},\ u_{1,qhk},-\rho_{ebeo}\right)\right)-\right.$$

$$\left(\frac{-u_{1,qh}u_{1,qhk-1}+\rho_{ebeo}-\rho_{ebeo}\frac{u^2{}_{1,qh}+u^2{}_{1,qhk-1}+2\rho_{ebeo}u_{1,qh}u_{1,qhk-1}}{1-\rho^2{}_{ebeo}}}{1-\rho^2{}_{ebeo}}\varphi\left(u_{1,qh},\ u_{1,qhk-1},-\rho_{ebeo}\right)\right)$$

$$]-\left[\frac{u_{1,qh}u_{1,qhK}+\rho_{ebeo}-\rho_{ebeo}\frac{u^2{}_{1,qh}+u^2{}_{1,qhK}-2\rho_{ebeo}u_{1,qh}u_{1,qhK}}{1-\rho^2{}_{ebeo}}}{1-\rho^2{}_{ebeo}}\varphi\left(u_{1,qh},\ u_{1,qhK},\ \rho_{ebeo}\right)\right]\}$$

$$\left(\frac{-4\exp(2r_e)}{(1+\exp(2r_e))^2}\right)\left(\frac{4\exp(2r_e)}{(1+\exp(2r_e))^2}\right)+\{\left[\varphi\left(u_{1,qh},u_{1,qh0},-\rho_{ebeo}\right)\right]$$

$$+\left[\varphi\left(u_{1,qh},u_{1,qhk},-\rho_{ebeo}\right)-\varphi\left(u_{1,qh},u_{1,qhk-1},-\rho_{ebeo}\right)\right]$$

$$+\left[-\varphi\left(u_{1,qh},u_{1,qhK},\rho_{ebeo}\right)\right]\}\left(-\frac{8\exp(2r)-8\exp(4r)}{(1+\exp(2r))^3}\right)$$

$$-\frac{\partial \log f}{\partial r_e}\frac{\partial \log f}{\partial r_e} \qquad (31)$$

The second differentiation of (4.65) with respect to r is given by,

$$\frac{\partial^2 \log f}{\partial r \partial r_e} = \sum_{t=1}^{T} \sum_{k=0}^{K} d_{itk} \frac{1}{C} \{$$

$$\left[-a_q \sigma_{\delta_o} u_{1,qh0} \varphi\left(u_{1,qh0}, u_{1,qh}, -\rho_{ebeo}\right) - \left(-a_h \sigma_{\delta_b} + a_q \sigma_{\delta_o} \rho_{ebeo}\right)\right.$$

$$\left. - \left(-a_h \sigma_{\delta_b} + a_q \sigma_{\delta_o} \rho_{ebeo}\right) \frac{u_{1,qh} + \rho_{ebeo} u_{1,qh0}}{1 - \rho_{ebeo}^2} \varphi\left(u_{1,qh0}, u_{1,qh}, -\rho_{ebeo}\right)\right]$$

$$\left[\left(-a_q \sigma_{\delta_o} u_{1,qhk} \varphi\left(u_{1,qhk}, u_{1,qh}, -\rho_{ebeo}\right) - \left(-a_h \sigma_{\delta_b} + a_q \sigma_{\delta_o} \rho_{ebeo}\right)\right.\right.$$

$$\left.\frac{u_{1,qh} + \rho_{ebeo} u_{1,qhk}}{1 - \rho_{ebeo}^2} \varphi\left(u_{1,qhk}, u_{1,qh}, -\rho_{ebeo}\right)\right) - \left(-a_q \sigma_{\delta_o} u_{1,qhk-1} \varphi\left(u_{1,qhk-1}, u_{1,qh}, -\rho_{ebeo}\right)\right.$$

$$\left. - \left(-a_h \sigma_{\delta_b} + a_q \sigma_{\delta_o} \rho_{ebeo}\right) \frac{u_{1,qh} + \rho_{ebeo} u_{1,qhk-1}}{1 - \rho_{ebeo}^2} \varphi\left(u_{1,qhk-1}, u_{1,qh}, -\rho_{ebeo}\right)\right)$$

$$\left[-a_q \sigma_{\delta_o} u_{1,qhK} \varphi\left(u_{1,qhK}, u_{1,qh}, -\rho_{ebeo}\right) + \left(-a_h \sigma_{\delta_b} + a_q \sigma_{\delta_o} \rho_{ebeo}\right)\right.$$

$$\frac{\sqrt{2} \rho_{\delta_{1i} \delta_{2i}}}{\sqrt{\left(1 - \rho_{\delta_{1i} \delta_{2i}}^2\right)}} \frac{-4 \exp(2r_e)}{(1 + \exp(2r_e))^2} \frac{4 \exp(2r)}{(1 + \exp(2r))^2}$$

$$\frac{4 \exp(2r)}{(1 + \exp(2r))^2} \frac{-4 \exp(2r_e)}{(1 + \exp(2r_e))^2} \frac{4 \exp(2r)}{(1 + \exp(2r))^2} - \frac{\partial \log f}{\partial r} \frac{\partial \log f}{\partial r_e} \tag{32}$$

The second differentiation of (4.65) with respect to α_1 is given by,

$$\frac{\partial^2 \log f}{\partial \alpha_1 \partial r_e} = \sum_{t=1}^{T} \sum_{k=0}^{K} d_{itk} \frac{1}{C} \{$$

$$\left[-u_{1,qh1} \varphi\left(u_{1,qh1}, u_{1,qh}, -\rho_{ebeo}\right) - \frac{\rho_{ebeo}\left(u_{1,qh} + \rho_{ebeo} u_{1,qh1}\right)}{1 - \rho_{ebeo}^2} \varphi\left(u_{1,qh1}, u_{1,qh}, -\rho_{ebeo}\right)\right]$$

$$- \left[-u_{1,qh1} \varphi\left(u_{1,qh1}, u_{1,qh}, -\rho_{ebeo}\right) - \frac{\rho_{ebeo}\left(u_{1,qh} + \rho_{ebeo} u_{1,qh1}\right)}{1 - \rho_{ebeo}^2} \varphi\left(u_{1,qh1}, u_{1,qh}, -\rho_{ebeo}\right)\right]$$

$$\} \frac{-4 \exp(2r_e)}{(1 + \exp(2r_e))^2} - \frac{\partial \log f}{\partial \alpha_1} \frac{\partial \log f}{\partial r_e} \tag{33}$$

The second differentiation of (4.65) with respect toα_k is given by,

$$\frac{\partial^2 \log f}{\alpha_k \partial r_e} = \sum_{t=1}^{T} \sum_{k=0}^{K} d_{itk} \frac{1}{C} \{$$

$$\left[-u_{1,qh1} \varphi \left(u_{1,qh1}, u_{1,qh}, -\rho_{ebeo} \right) - \frac{\rho_{ebeo} \left(u_{1,qh} + \rho_{ebeo} u_{1,qh1} \right)}{1 - \rho_{ebeo}^2} \varphi \left(u_{1,qh1}, u_{1,qh}, -\rho_{ebeo} \right) \right] -$$

$$\left[\left[-u_{1,qh1} \varphi \left(u_{1,qh1}, u_{1,qh}, -\rho_{ebeo} \right) - \frac{\rho_{ebeo} \left(u_{1,qh} + \rho_{ebeo} u_{1,qh1} \right)}{1 - \rho_{ebeo}^2} \varphi \left(u_{1,qh1}, u_{1,qh}, -\rho_{ebeo} \right) \right] \right\}$$

$$\frac{-4 \exp(2r_e)}{(1 + \exp(2r_e))^2} - \frac{\partial \log f}{\partial \alpha_k} \frac{\partial \log f}{\partial r_e} \tag{34}$$

The second differentiation of (4.66) with respect to r is given by,

$$\frac{\partial^2 \log f}{\partial r^2} = \left(\frac{-\left(1+\rho^2_{\delta_{1i}\delta_{2i}}\right)}{\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)^2} \right) \left(\frac{4\exp(2r)}{(1+\exp(2r))^2} \right)^2$$
$$-\frac{\rho_{\delta_{1i}\delta_{2i}}}{\left(1-\rho^2_{\delta_{1i}\delta_{2i}}\right)} \frac{8\exp(2r) - 8\exp(4r)}{(1+\exp(2r))^3} + 2a_q a_h \frac{8\exp(2r) - 8\exp(4r)}{(1+\exp(2r))^3}$$
$$+\sum_{t=1}^{T}\sum_{k=0}^{K} d_{itk}\frac{1}{C}\left\{\left[u_{1,qh}\varphi\left(u_{1,qh}\right)a_h^2\sigma^2_{\delta_b} + \right.\right.$$
$$\left(-u_{1,qh}\varphi\left(u_{1,qh}\right)a_h^2\sigma^2_{\delta_b}\Phi\left(\frac{u_{1,qh0}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho^2_{ebeo}}}\right) + a_h\sigma_{\delta_b}\left(a_h\sigma_{\delta_1}\rho_{ebeo} - a_q\sigma_{\delta_2}\right)\right.$$
$$\varphi\left(u_{1,qh0},u_{1,qh},-\rho_{ebeo}\right) - u_{1,qh0}\varphi\left(u_{1,qh0}\right)a_q^2\sigma_2^2\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qh0}}{\sqrt{1-\rho^2_{ebeo}}}\right)$$
$$\left.\left. +a_q\sigma_{\delta_o}\left(-a_h\sigma_{\delta_1} + a_q\sigma_{\delta_2}\rho_{ebeo}\right)\varphi\left(u_{1,qh0},u_{1,qh},-\rho_{ebeo}\right)\right)\right] +$$
$$\left(-u_{1,qh}\varphi\left(u_{1,qh}\right)a_h^2\sigma^2_{\delta_b}\Phi\left(\frac{u_{1,qhk}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho^2_{ebeo}}}\right) + a_h\sigma_{\delta_b}\left(a_h\sigma_{\delta_1}\rho_{ebeo} - a_q\sigma_{\delta_2}\right)\right.$$
$$\varphi\left(u_{1,qhk},u_{1,qh},-\rho_{ebeo}\right) - u_{1,qhk}\varphi\left(u_{1,qhk}\right)a_q^2\sigma_2^2\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk}}{\sqrt{1-\rho^2_{ebeo}}}\right)$$
$$\left. +a_q\sigma_{\delta_o}\left(-a_h\sigma_{\delta_1} + a_q\sigma_{\delta_2}\rho_{ebeo}\right)\varphi\left(u_{1,qhk},u_{1,qh},-\rho_{ebeo}\right)\right) -$$
$$-u_{1,qh}\varphi\left(u_{1,qh}\right)a_h^2\sigma^2_{\delta_b}\Phi\left(\frac{u_{1,qhk-1}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho^2_{ebeo}}}\right) + a_h\sigma_{\delta_b}\left(a_h\sigma_{\delta_1}\rho_{ebeo} - a_q\sigma_{\delta_2}\right)$$
$$\varphi\left(u_{1,qhk-1},u_{1,qh},-\rho_{ebeo}\right) - u_{1,qhk}\varphi\left(u_{1,qhk}\right)a_q^2\sigma_2^2\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk-1}}{\sqrt{1-\rho^2_{ebeo}}}\right)$$
$$+a_q\sigma_{\delta_o}\left(-a_h\sigma_{\delta_1} + -a_q\sigma_{\delta_2}\rho_{ebeo}\right)\varphi\left(u_{1,qhk-1},u_{1,qh},-\rho_{ebeo}\right)\left.\right)\left.\right]$$
$$\left[\left(-u_{1,qh}\varphi\left(u_{1,qh}\right)a_h^2\sigma^2_{\delta_b}\Phi\left(\frac{u_{1,qhK}-\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho^2_{ebeo}}}\right) + a_h\sigma_{\delta_b}\left(a_q\sigma_{\delta_o} - \rho_{ebeo}a_h\sigma_{\delta_b}\right)\right.\right.$$
$$\varphi\left(u_{1,qh},u_{1,qhK},\rho_{ebeo}\right) - u_{1,qhK}\varphi\left(u_{1,qhK}\right)a_q^2\sigma_2^2\Phi\left(\frac{u_{1,qh}-\rho_{ebeo}u_{1,qhK}}{\sqrt{1-\rho^2_{ebeo}}}\right)$$
$$\left.\left. +a_q\sigma_{\delta_o}\left(a_h\sigma_{\delta_b} - \rho_{ebeo}a_q\sigma_{\delta_o}\right)\varphi\left(u_{1,qh},u_{1,qhK},\rho_{ebeo}\right)\right)\right]$$
$$\frac{\sqrt{2}\rho_{\delta_{1i}\delta_{2i}}}{\sqrt{1-\rho^2_{\delta_{1i}\delta_{2i}}}}\frac{\sqrt{2}\rho_{\delta_{1i}\delta_{2i}}}{\sqrt{1-\rho^2_{\delta_{1i}\delta_{2i}}}}\left(\frac{4\exp(2r)}{(1+\exp(2r))^2}\right)^2 + \tag{35}$$

$$\left\{\left[\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}+\right.\right. .$$

$$\left(-\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}\Phi\left(\frac{u_{1,qh0}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)+\varphi\left(u_{1,qh0}\right)a_q\sigma_{\delta_o}\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qh0}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right)\Bigg]$$

$$\left[\left(-\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}\Phi\left(\frac{u_{1,qhk}+\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)+\varphi\left(u_{1,qhk}\right)a_q\sigma_{\delta_o}\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qhk}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right)-\right.$$

$$\left.\left.\left[-\varphi\left(u_{1,qh}\right)a_h\sigma_{\delta_b}\Phi\left(\frac{u_{1,qhK}-\rho_{ebeo}u_{1,qh}}{\sqrt{1-\rho_{ebeo}^2}}\right)-\varphi\left(u_{1,qhK}\right)a_q\sigma_{\delta_o}\Phi\left(\frac{u_{1,qh}-\rho_{ebeo}u_{1,qhK}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right]\right]\right\}$$

$$\frac{\sqrt{2}\rho_{\delta_{1i}\delta_{2i}}}{\sqrt{1-\rho_{\delta_{1i}\delta_{2i}}^2}}\frac{8\exp(2r)-8\exp(4r)}{(1+\exp(2r))^3}-\frac{\partial\log f}{\partial r}\frac{\partial\log f}{\partial r} \quad (36)$$

The second differentiation of (4.66) with respect to α_1 is given by,

$$\frac{\partial^2\log f}{\partial\alpha_1\partial r}=\sum_{t=1}^{T}\sum_{k=0}^{K}d_{itk}\frac{1}{C}\left\{\left[-a_h\sigma_{\delta_b}\varphi\left(u_{1,qh1},\ u_{1,qh},-\rho_{ebeo}\right)-\right.\right.$$

$$\left.u_{1,qh1}\varphi\left(u_{1,qh1}\right)a_q\sigma_{\delta_o}\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qh1}}{\sqrt{1-\rho_{ebeo}^2}}\right)+a_q\sigma_{\delta_o}\rho_{ebeo}\varphi\left(u_{1,qh1},\ u_{1,qh},-\rho_{ebeo}\right)\right]$$

$$+\left[-\left(-a_h\sigma_{\delta_b}\varphi\left(u_{1,qh1},\ u_{1,qh},-\rho_{ebeo}\right)-u_{1,qh0}\varphi\left(u_{1,qh1}\right)a_q\sigma_{\delta_o}\Phi\left(\frac{u_{1,qh}+\rho_{ebeo}u_{1,qh1}}{\sqrt{1-\rho_{ebeo}^2}}\right)\right.\right.$$

$$\left.\left.+a_q\sigma_{\delta_o}\rho_{ebeo}\varphi\left(u_{1,qh1},\ u_{1,qh},-\rho_{ebeo}\right)\right)\right]\frac{\sqrt{2}\rho_{\delta_{1i}\delta_{2i}}}{\sqrt{\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}}\frac{4\exp(2r)}{(1+\exp(2r))^2}-\frac{\partial\log f}{\partial\alpha_1}\frac{\partial\log f}{\partial r}$$

$$(37)$$

The second differentiation of (4.66) with respect to α_k is given by,

$$\frac{\partial^2 \log f}{\partial \alpha_k \partial r} = \sum_{t=1}^{T} \sum_{k=0}^{K} d_{itk} \frac{1}{C} \left\{ \left[-a_h \sigma_{\delta_b} \varphi\left(u_{1,qhk},\ u_{1,qh}, -\rho_{ebeo}\right) - \right. \right.$$
$$\left. u_{1,qhk} \varphi\left(u_{1,qhk}\right) a_q \sigma_{\delta_o} \Phi\left(\frac{u_{1,qh} + \rho_{ebeo} u_{1,qhk}}{\sqrt{1-\rho_{ebeo}^2}}\right) + a_q \sigma_{\delta_o} \rho_{ebeo} \varphi\left(u_{1,qhk},\ u_{1,qh}, -\rho_{ebeo}\right) \right]$$
$$\left[- \left(a_h \sigma_{\delta_b} \varphi\left(u_{1,qhk},\ u_{1,qh}, -\rho_{ebeo}\right) - u_{1,qhk} \varphi\left(u_{1,qhk}\right) a_q \sigma_{\delta_o} \Phi\left(\frac{u_{1,qh} + \rho_{ebeo} u_{1,qhk}}{\sqrt{1-\rho_{ebeo}^2}}\right) \right. \right.$$
$$\left. \left. -a_q \sigma_{\delta_o} \rho_{ebeo} \varphi\left(u_{1,qhk},\ u_{1,qh}, -\rho_{ebeo}\right) \right) \right] \frac{\sqrt{2} \rho_{\delta_{1i}\delta_{2i}}}{\sqrt{\left(1-\rho_{\delta_{1i}\delta_{2i}}^2\right)}} \frac{4\exp(2r)}{(1+\exp(2r))^2} - \frac{\partial \log f}{\partial \alpha_k} \frac{\partial \log f}{\partial r}$$

(38)

The second differentiation of (4.68) with respect to α_1 is given by,

$$\frac{\partial^2 \log f}{\partial \alpha_1 \partial \alpha_1} = \sum_{t=1}^{T} \sum_{k=0}^{K} d_{itk} \frac{1}{C} \{$$
$$\left[-u_{1,qh1} \varphi\left(u_{1,qh1}\right) \Phi\left(\frac{u_{1,qh} + \rho_{ebeo} u_{1,qh1}}{\sqrt{1-\rho^2{}_{ebeo}}}\right) + \rho_{ebeo} \varphi\left(u_{1,qh}, u_{1,qh1}, -\rho_{ebeo}\right) \right] +$$
$$\left[- \left(-u_{1,qh1} \varphi\left(u_{1,qh1}\right) \Phi\left(\frac{u_{1,qh} + \rho_{ebeo} u_{1,qh1}}{\sqrt{1-\rho^2{}_{ebeo}}}\right) + \rho_{ebeo} \varphi\left(u_{1,qh}, u_{1,qh1}, -\rho_{ebeo}\right) \right) \right] \}$$
$$- \frac{\partial \log f}{\partial \alpha_1} \frac{\partial \log f}{\partial \alpha_1} \tag{39}$$

The second differentiation of (4.68) with respect to α_k is given by,

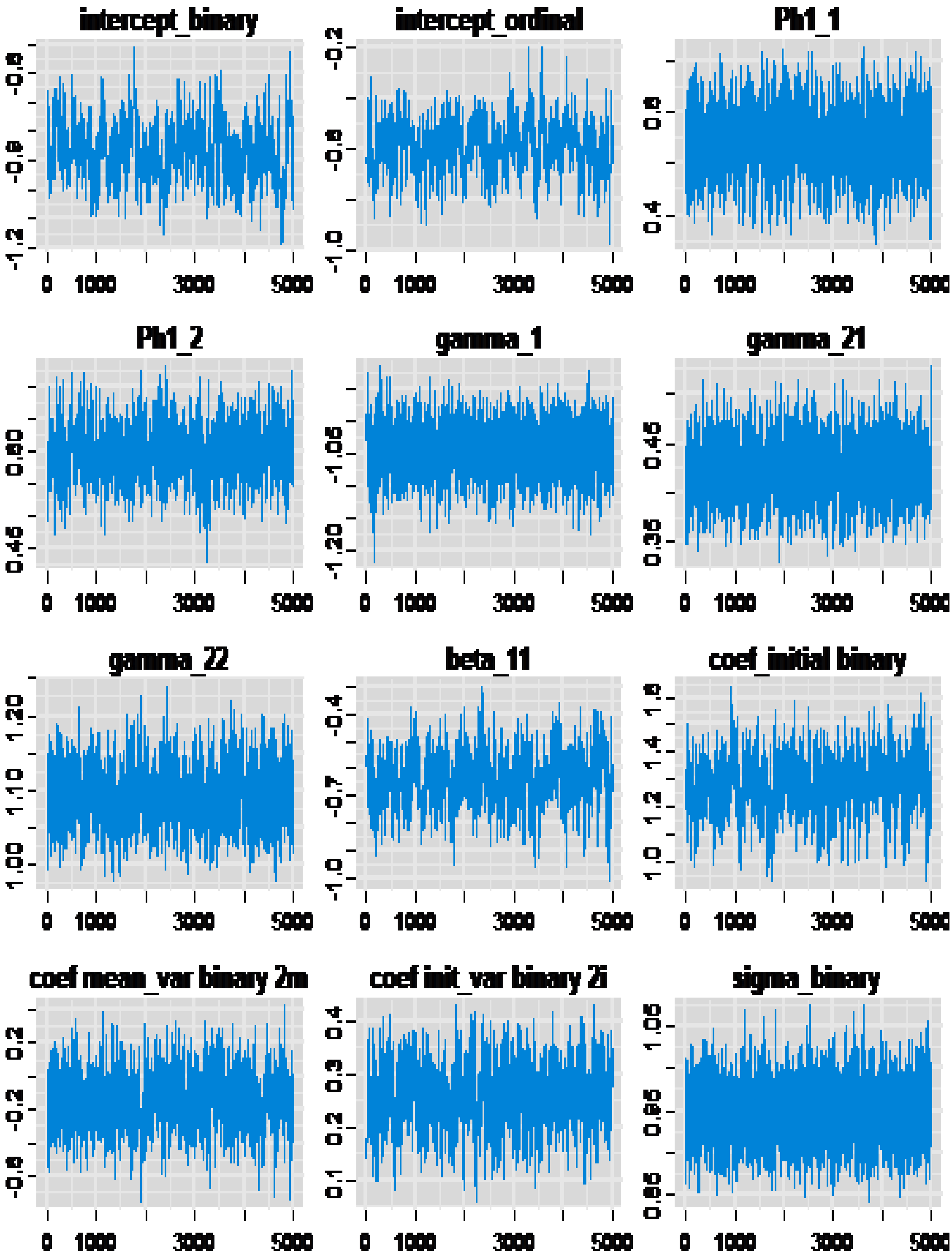
$$\frac{\partial^2 \log f}{\partial \alpha_k \partial \alpha_1} = \sum_{t=1}^{T} \sum_{k=0}^{K} d_{itk} \frac{1}{C} (0) - \frac{\partial \log f}{\partial \alpha_k} \frac{\partial \log f}{\partial \alpha_1} \tag{40}$$

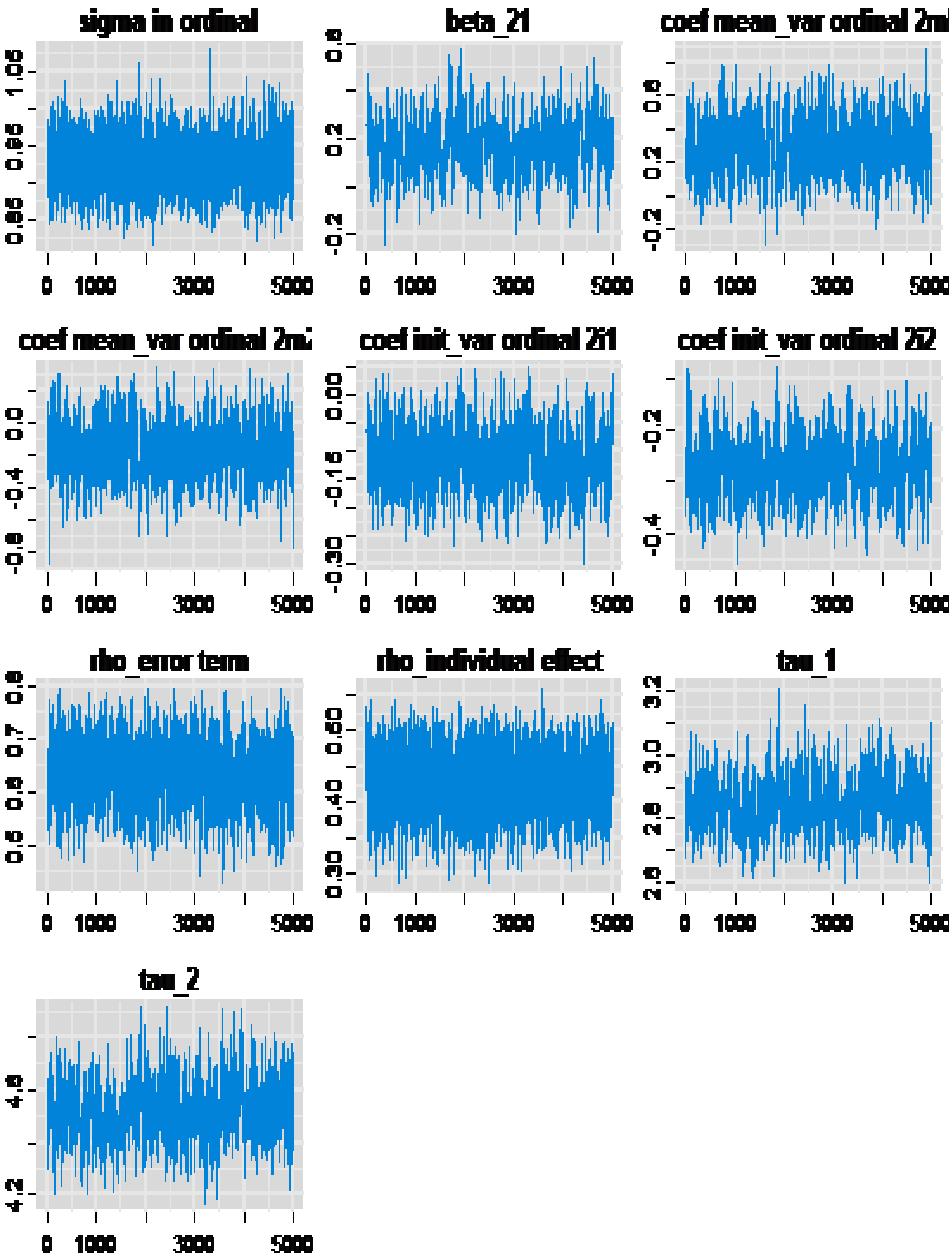
The second differentiation of (4.69) with respect to α_k is given by,

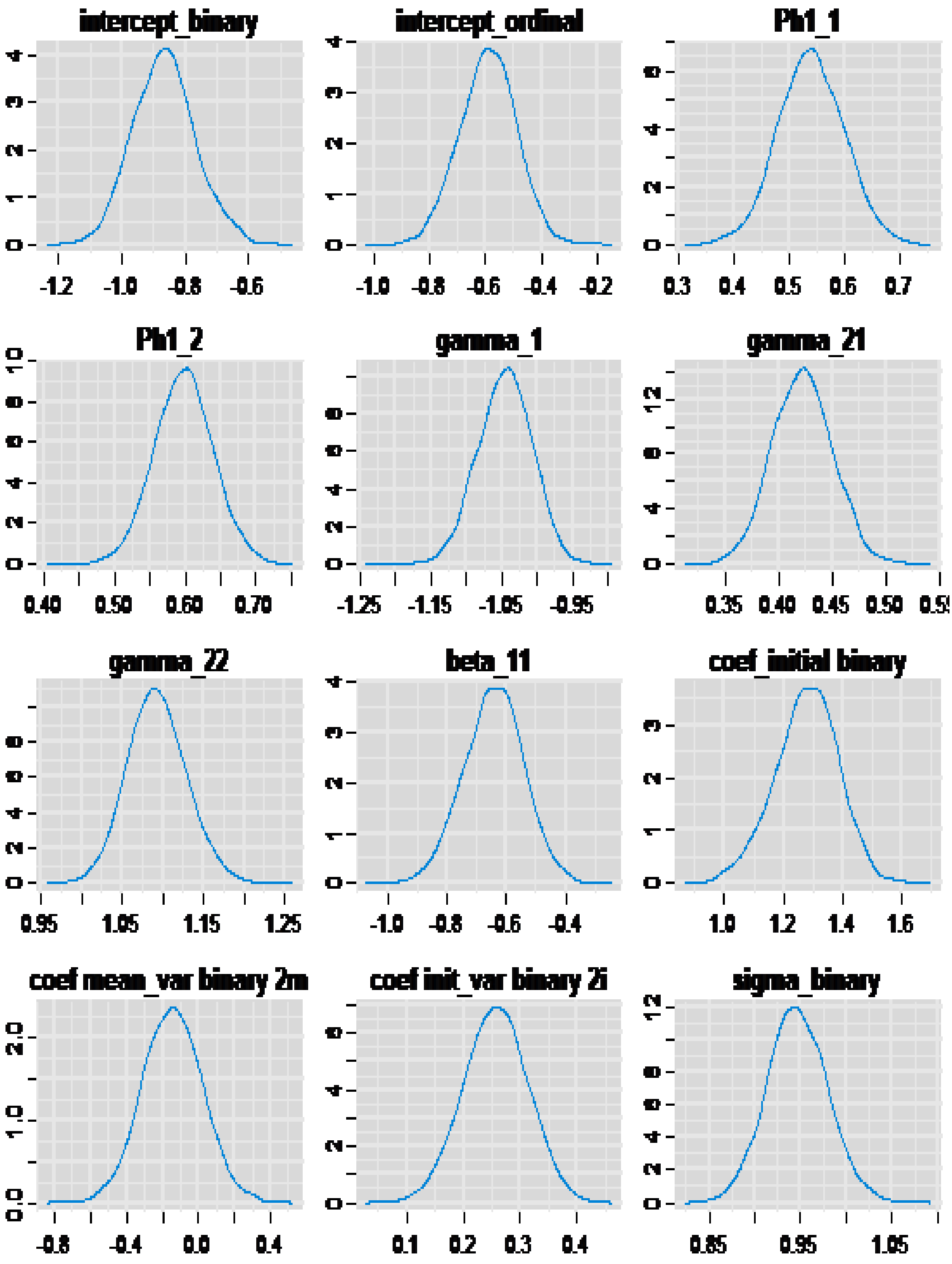
$$\frac{\partial^2 \log f}{\partial \alpha_k \partial \alpha_k} = \sum_{t=1}^{T} \sum_{k=0}^{K} d_{itk} \frac{1}{C} \{$$
$$\left[-\varphi\left(u_{1,qhk}\right) \Phi\left(\frac{u_{1,qh} + \rho_{ebeo} u_{1,qhk}}{\sqrt{1-\rho^2{}_{ebeo}}} \right) + \rho_{ebeo} \varphi\left(u_{1,qh}, u_{1,qhk}, -\rho_{ebeo}\right) \right]$$
$$\left[- \left(-\varphi\left(u_{1,qhk}\right) \Phi\left(\frac{u_{1,qh} + \rho_{ebeo} u_{1,qhk}}{\sqrt{1-\rho^2{}_{ebeo}}} \right) + \rho_{ebeo} \varphi\left(u_{1,qh}, u_{1,qhk}, -\rho_{ebeo}\right) \right) \right] \Bigg\} (\mathrm{e}^{\alpha_k})^2$$
$$- \frac{\partial \log f}{\partial \alpha_k} \frac{\partial \log f}{\partial \alpha_k} \tag{41}$$

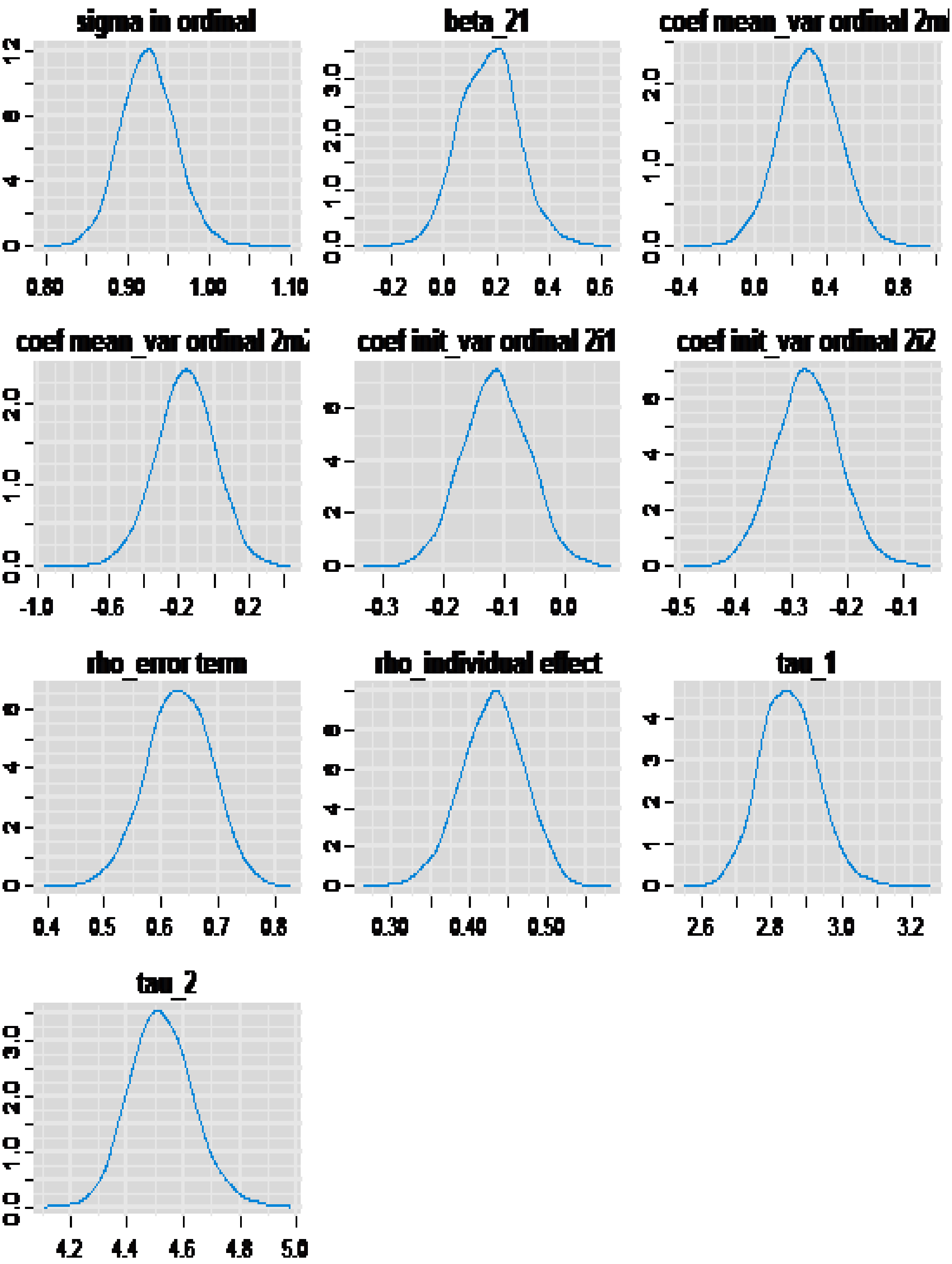
II Trace Plots and Densities for the Simulated Data

Figure 1: Trace Plots and Densities for the Simulated Data




sigma in ordinal
beta_21
coef mean_var ordinal 2m
coef mean_var ordinal 2m
coef init_var ordinal 2|1
coef init_var ordinal 2|2
rho_error term
rho_individual effect
tau_1
tau_2

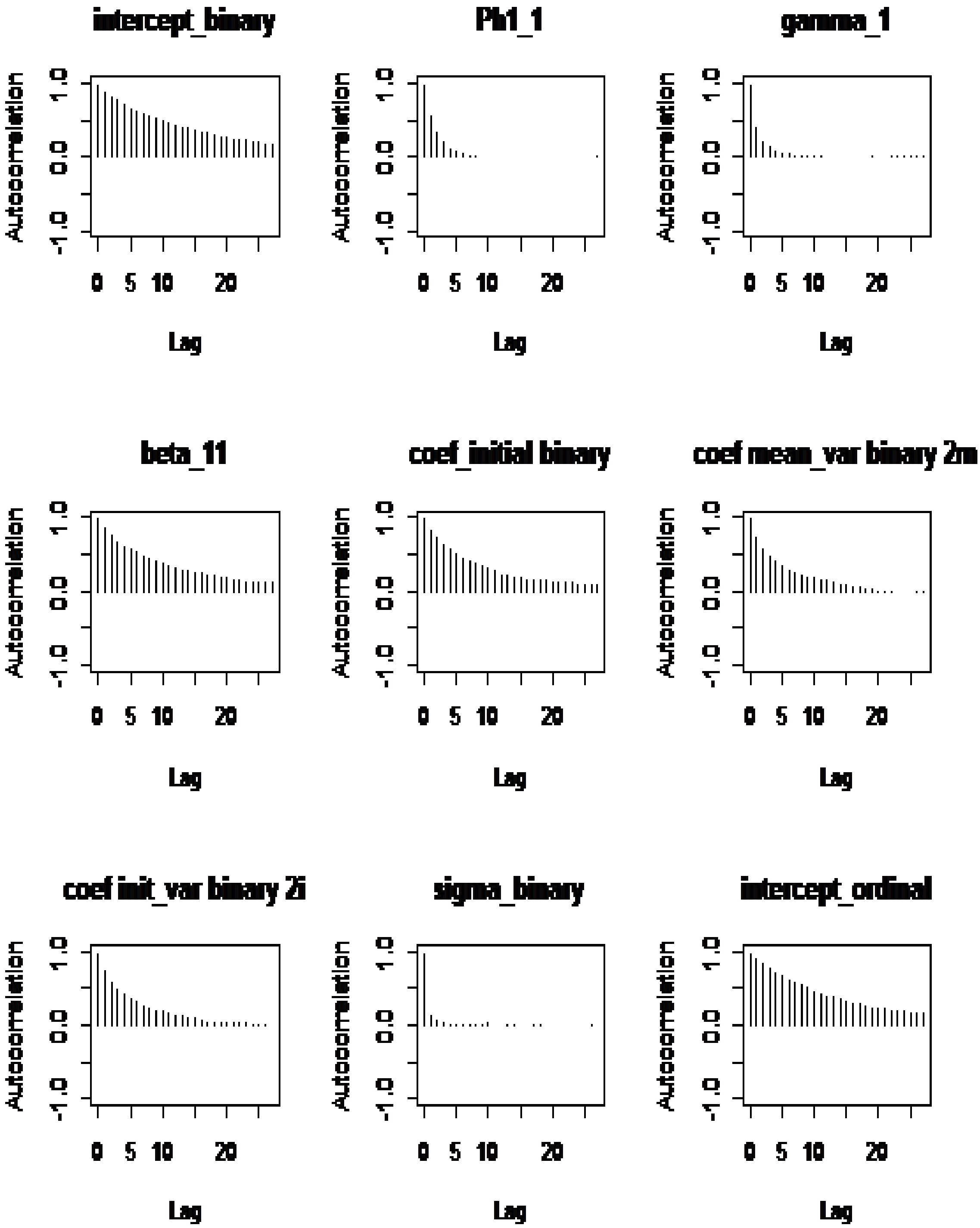

intercept_binary
intercept_ordinal
Ph1_1
Ph1_2
gamma_1
gamma_21
gamma_22
beta_11
coef_initial binary
coef mean_var binary 2m
coef init_var binary 2i
sigma_binary

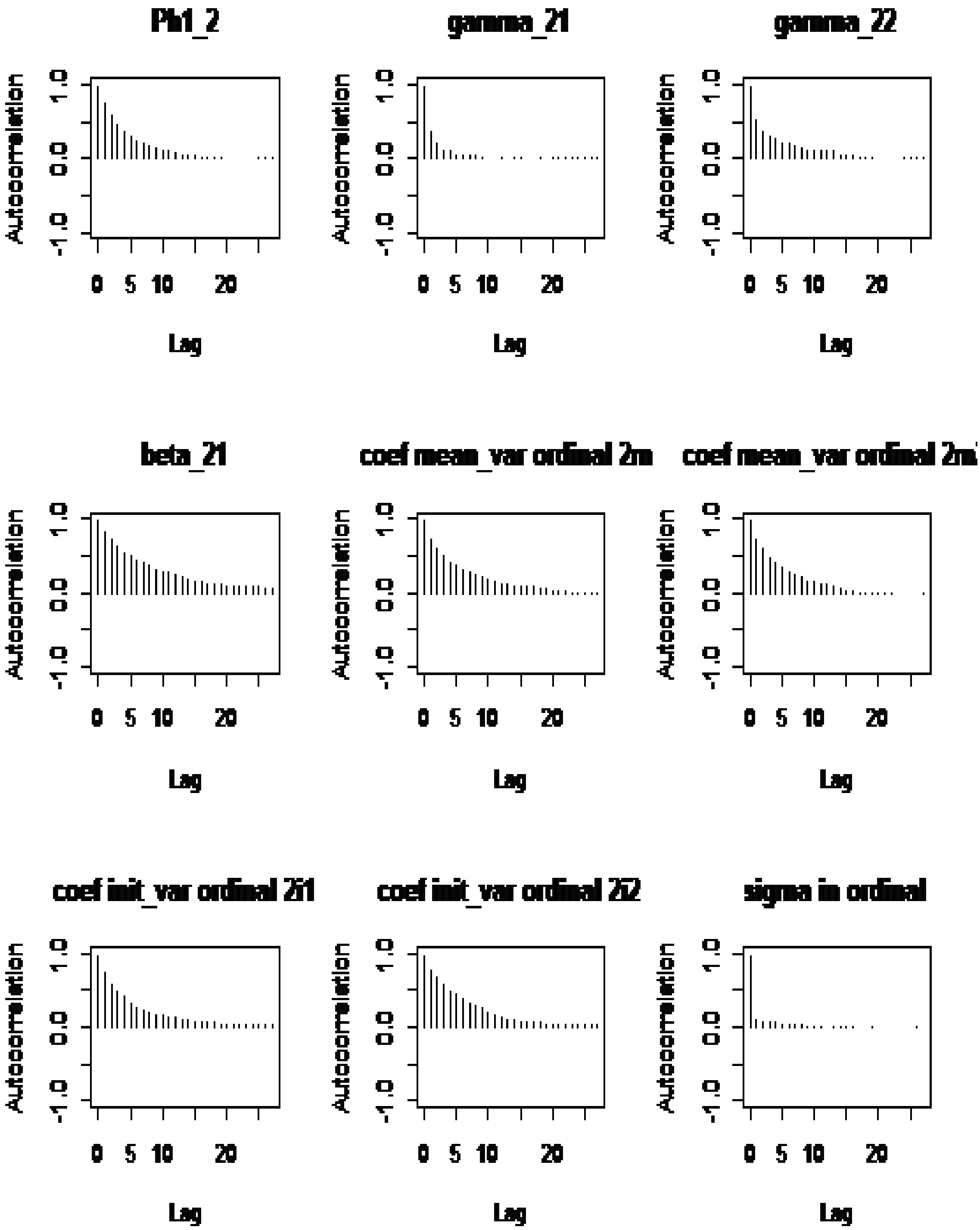


The trace plots and their densities are shown above. The diagnostic plots shows that generally, MCMC chains had a sensible convergence irrespective of the complexity of the model.

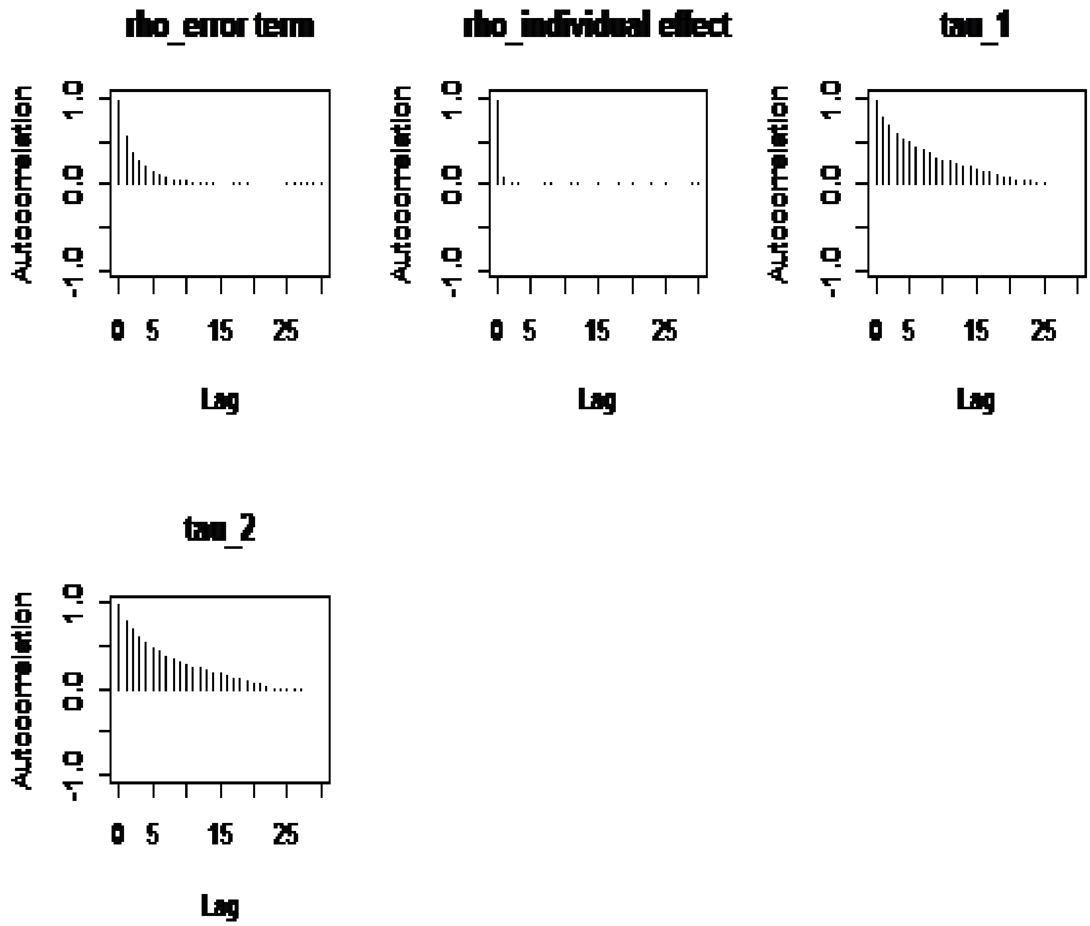
III Autocorrelation Plot for Simulated Data

Figure 2: Autocorrelation Plot for Simulated Data





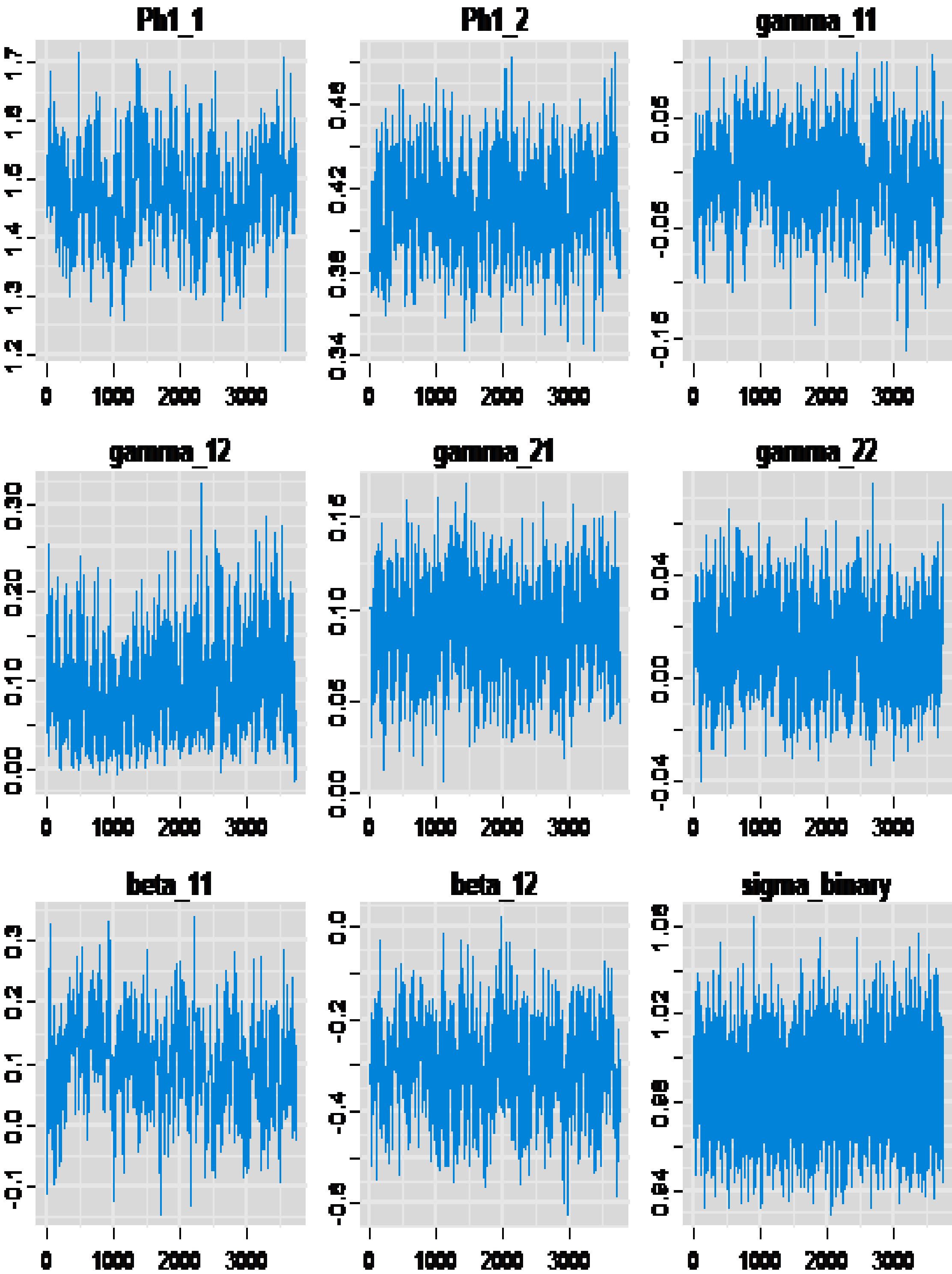
Phi1_2
gamma_21
gamma_22
beta_21
coef mean_var ordinal 2m
coef mean_var ordinal 2m
coef init_var ordinal 2i1
coef init_var ordinal 2i2
sigma in ordinal
Autocorrelation
1.0
0.0
-1.0
0 5 10 20
Lag

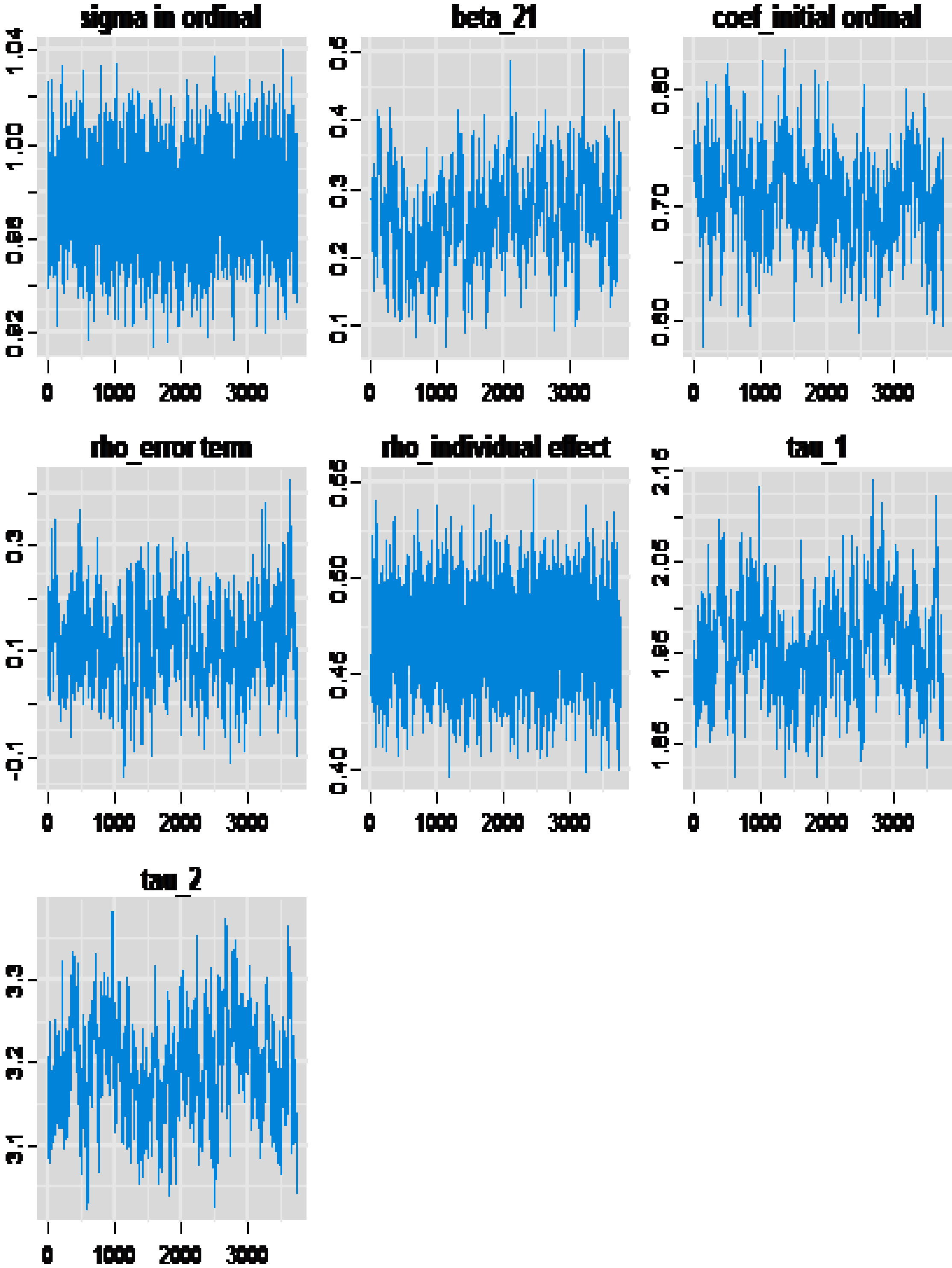


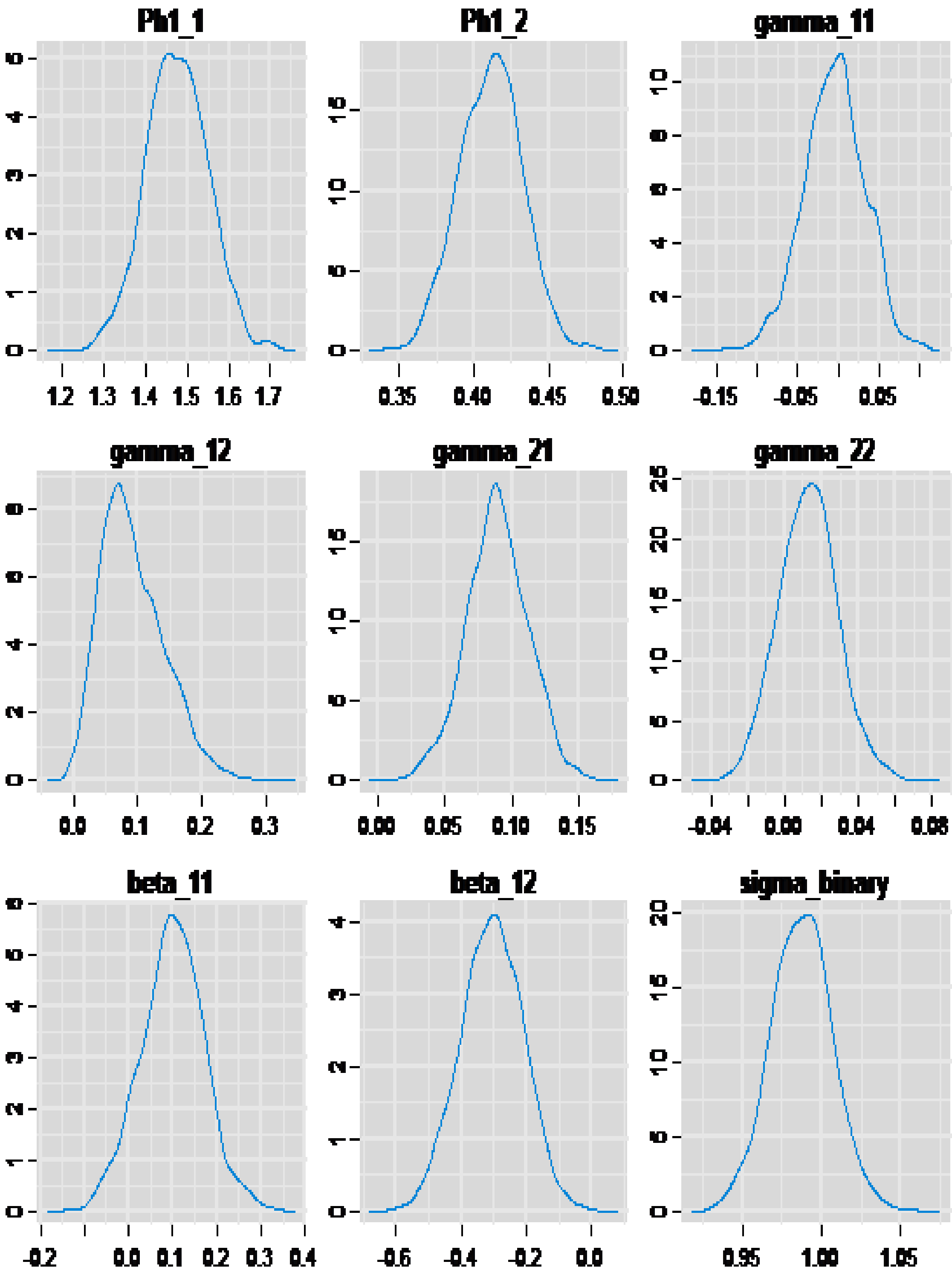
Initial screening of autocorrelation plots indicated a high autocorrelations between the generated values. After thinning, picked every 6th iteration after burning, autocorrelation between samples were considerably low to provide the independent samples to compute final statistics such as mean and standard errors.

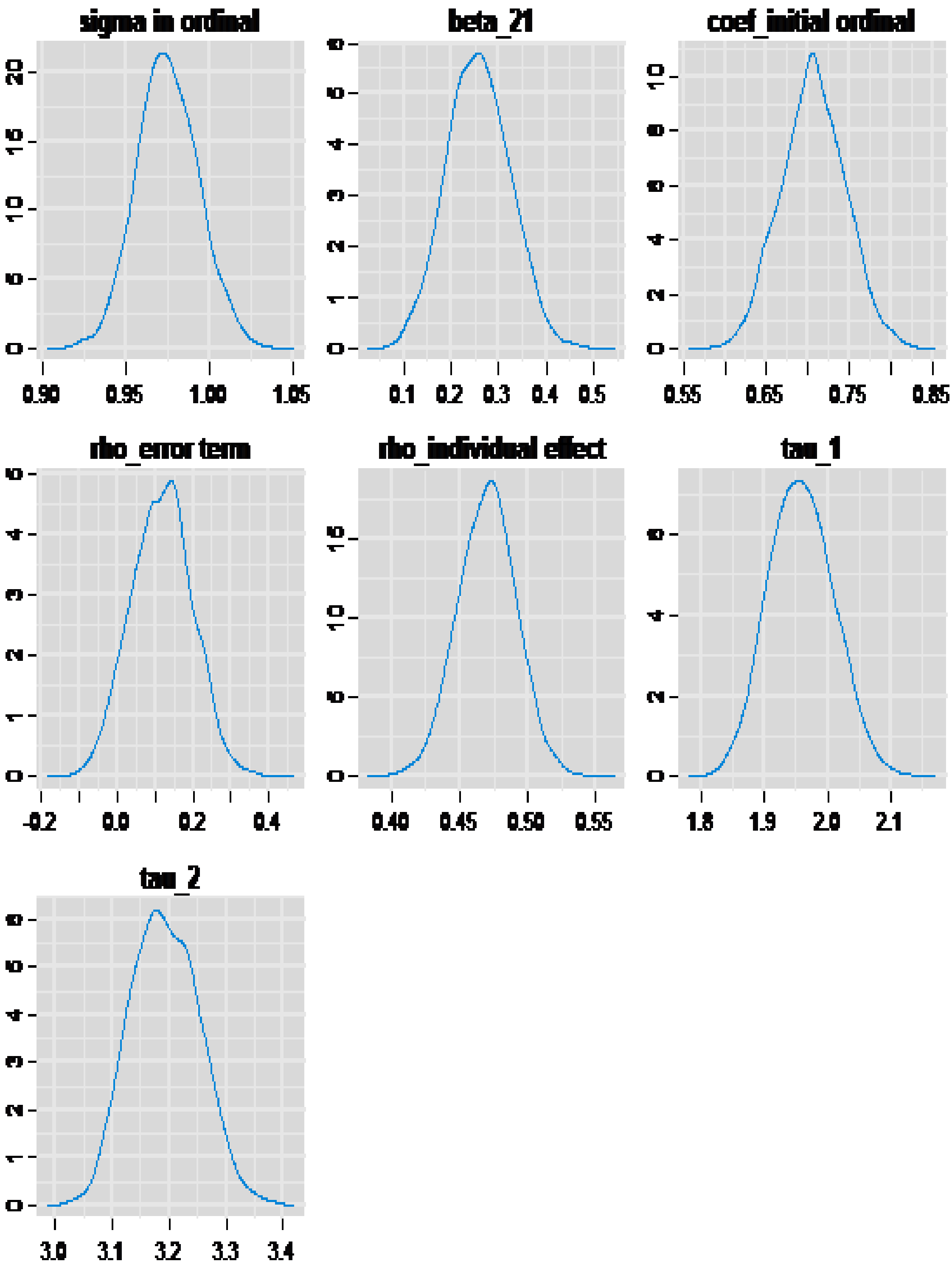
IV Trace Plots and Densities for the Smoking Data

Figure 3: Trace Plots and Densities for the Smoking Data




sigma in ordinal
beta_21
coef_initial ordinal
rho_error term
rho_individual effect
tau_1
tau_2
0
1000
2000
3000

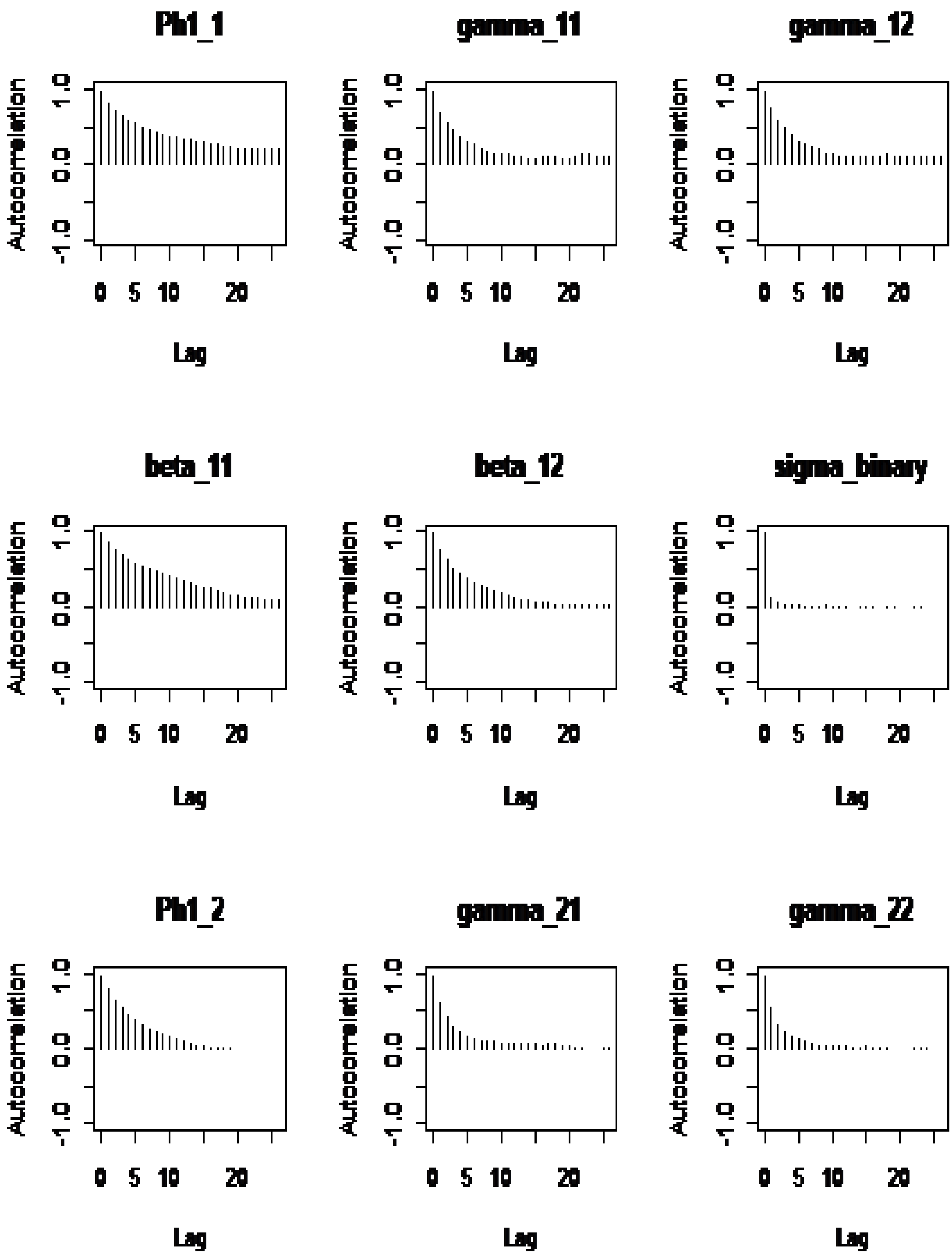

Phi1_1
Phi1_2
gamma_11
gamma_12
gamma_21
gamma_22
beta_11
beta_12
sigma_binary

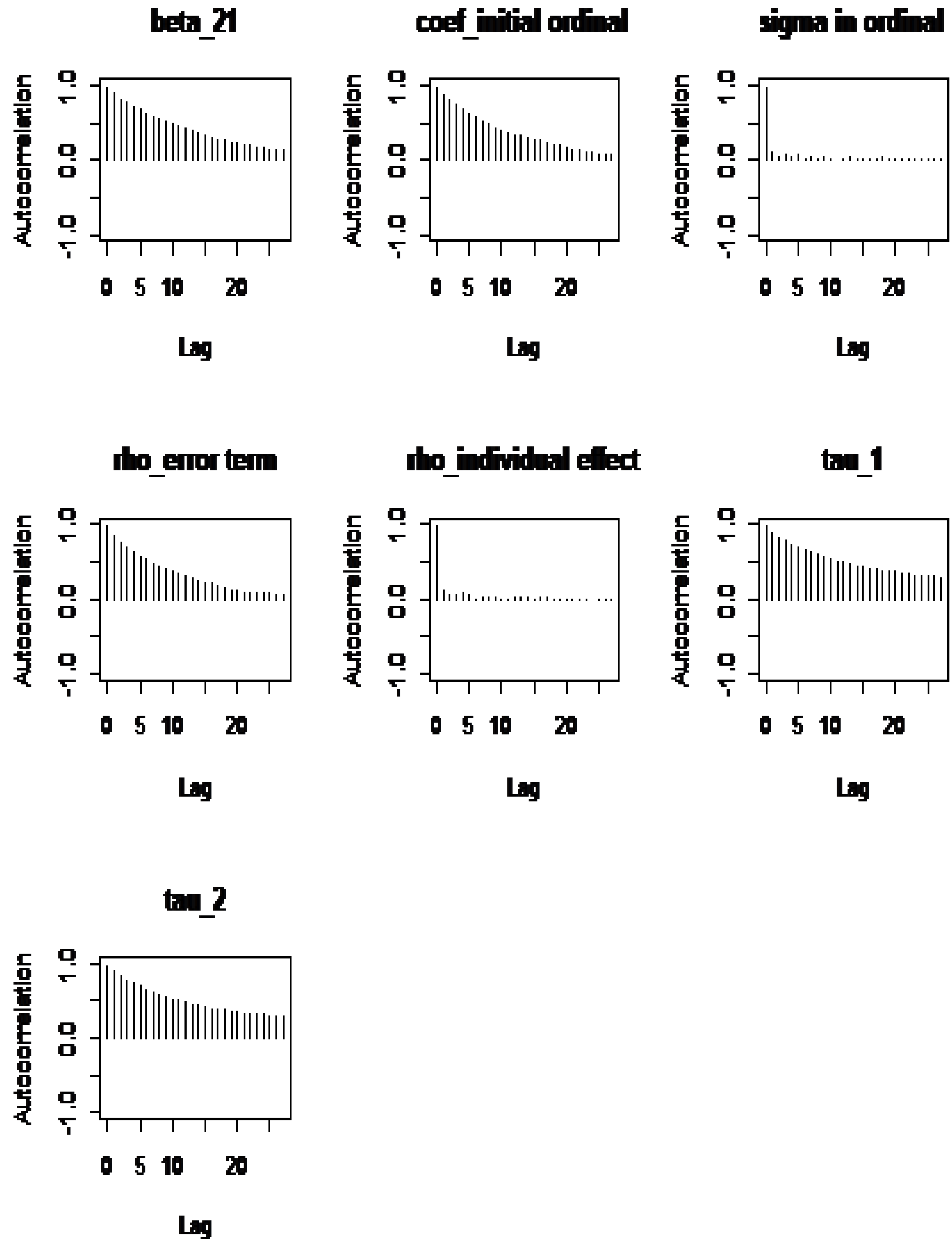


The trace plots and their densities are shown above. The diagnostic plots shows that generally, MCMC chains had a sensible convergence irrespective of the complexity of the model.

V Autocorrelation Plot for Smoking Data

Figure 4: Autocorrelation Plot for Smoking Data





Initial screening of autocorrelation plots indicated a high autocorrelations between the generated values. After thinning, picked every 6th iteration after burning, autocorrelation between samples were considerably low to provide the independent samples to compute final statistics such as mean and standard errors.

VI Articles Published from the Thesis

1. Kung'u, J., Odongo, L. and Kube, A. (2022) 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, 58-74. doi: 10.11648/j.ajtas.20221102.11

2. Kung'u, J., Kube, A. and Odongo, L. (2022) Bayesian Approach to Zero-Inflated Dynamic Panel Ordered Probit Model with an Application in Drug Abuse. International Journal of Statistics and Applied Mathematics, 7(2): 01-13