This is to certify that the

dissertation entitled

Dynamic Unobserved Effects Model
for Continuous and Binary Response

presented by

Chung-Jung Lee

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Economics

[signature]
Major professor

Date 10/26/00

Dynamic Unobserved Effects Model for Continuous and Binary Response

By

Chung-Jung Lee

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

2000

ABSTRACT

Dynamic Unobserved Effects Model for Continuous and Binary Response

By

Chung-Jung Lee

In this thesis I consider estimation of dynamic, unobserved effects panel data models for both continuous and discrete outcomes. In order to handle correlation between the unobserved heterogeneity and the initial condition, I use the method of conditional maximum likelihood estimation (CMLE). This method turns out to be tractable for nonlinear binary response models as well as for dynamic linear models when the unobserved heterogeneity interacts with the lagged dependent variable. The CMLE performs well compared with various competitors that have been proposed in the literature.

The thesis is in four chapters. Chapter 1 surveys the existing literature for estimating dynamic linear models with unobserved effects, with attention to various assumptions that have been made on the initial conditions. In Chapter 2 I study the CMLE for the linear, dynamic model with an additive unobserved effect. I show how to construct the conditional likelihood function-- which uses an assumption about the distribution of the unobserved effect given exogenous variables and the initial condition. Monte Carlo evidence is provided, with and without the normality assumption

in the conditional distribution for heterogeneity, and I include an empirical application to wage dynamics for employed men.

Chapter 3 considers the CMLE for a useful extension of the basic linear model. Namely, I allow the unobserved heterogeneity to interact with the lagged dependent variable. Apparently, this model has not been treated in the literature. Conditional MLE is especially useful for obtaining consistent estimators. Chapter 4 studies the dynamic logit model with an unobserved effect. Even in this case, where the conditional mean function is nonlinear, the CMLE is feasible and produces interesting results in an application to union membership.

To my father, Fu-Li Lee and my mother, Yue-Xia Lee-Wu

ACKNOWLEDGMENTS

It is impossible to express adequately my gratitude for the assistance and encouragement which many persons have given me in the preparation of this thesis. The first person to whom I must much owe is my thesis advisor, Professor Jeffrey M. Wooldridge. For his timely recommendation and continued instruction about this thesis, it can smoothly come out at last. Especially for the original idea about the construction of the dynamic panel model he came up with and numerous associated suggestions and valuable advice, I, here, want to acknowledge special debt to my thesis advisor. I also thank the other members of my committee: Professors de Jong and Strauss. I must thank my parents and the rest of my family members, Chung-yi, Chung-chuan, Chung-Liang, Li-juan and Li-mei. Without their continual support and encouragement, I cannot go though with the Ph.D program. I also acknowledge my intimate friends, Chang-shun, Rei-yuan, Feng-yue, Hong-yao and shang-yu for providing the software and timely PC maintenance during the preparation of the thesis. Last, not the least, there are still a lot of persons who help me behind, but I have not mentioned in the list. I always appreciate all of their help with gratitude.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

Overview Of The Linear AR(1) Model With Unobserved Effects

1.1 Introduction

The AR(1) panel data model with an additive, unobserved effect has received much attention in recent years. Nickell (1981) noted that the usual within, or fixed effects, estimator was inconsistent with fixed time series dimension (T) as the cross section dimension (N) gets large. Further, maximum likelihood approaches that either treat the initial condition as nonrandom – in particular, independent of the unobserved heterogeneity – are also inconsistent for fixed T. As most panel data sets on individuals, families, and even firms are characterized by small T and large N, the interest in obtaining a consistent estimator of the autoregressive root with fixed T has become an important problem.

Anderson and Hsiao (1982) (AH for short) show how a simple instrument variables (IV) estimator, obtained from the first-difference equation, is consistent for fixed T. Subsequently, the AH estimator was shown to have poor properties when the autoregressive root is large (see Arellano and Bond [1991], Sevestre and Trognon

[1990]). More recent work has proposed additional moment conditions, often based on further assumptions, that can be used in generalized method of moments (GMM) estimation to improve upon the basic IV estimator(e.g. Arellano and Bond, [1991], Arellano and Bover [1995], and Ahn and Schmidt, [1995]). Furthermore, Blundell and Bond (1998) and Hahn (1999) recently show that the gain of efficiency of GMM over a certain range of parameter space is significant.

The current chapter is to give an overview of the prevailing estimators for a linear dynamic panel data model with an additive, unobserved effect. A commonly used dynamic model for panel data in the AR(1) model:

$$y_{it} = \rho\, y_{i,t-1} + x_{it}\beta_i + u_{it}, \quad i = 1, ..., N\ ;\ t = 1, ..., T, \tag{1.1}$$

where $u_{it} = a_i + \varepsilon_{it}$ and a_i is unobserved heterogeneity. Here, y_{it} is a scalar and x_{it} is a K-vector random variable. Most available panel data sets contain a large number of observations on individuals (N) over a limited number of periods (T), which means that a sensible asymptotic analysis treats $N \rightarrow \infty$ with fixed T. With T fixed, the stationarity assumption $\rho < 1$ is not necessary for usual inference procedures, but $\rho < 1$ is relevant for most of empirical applications. Because we just consider the case of $N \rightarrow \infty$ with fixed T, we do not distinguish semi-consistency with consistency and just adopt the consistency term (or inconsistency) standing for semi-consistency (or semi-inconsistency)(Nerlove and Balestra [1966]). We use the setup of (1.1) as a standard model to discuss different approaches to estimate the parameters in the following sections. Sometimes, we use full matrix notation to express equation (1.1) as follows:

$$Y = \rho Y_{-1} + X\underline{\beta} + Da + \varepsilon, \tag{1.2}$$

with

$$Y = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1T} \\ \vdots \\ y_{NT} \end{pmatrix}_{NT\times 1}, \quad Y_{-1} = \begin{pmatrix} y_{10} \\ \vdots \\ y_{1,T-1} \\ \vdots \\ y_{N,T-1} \end{pmatrix}_{NT\times 1}, \quad X = \begin{pmatrix} x_{11}^1 & \dots & x_{11}^k \\ \vdots & \ddots & \vdots \\ x_{1T}^1 & \dots & x_{1T}^k \\ \vdots & \ddots & \vdots \\ x_{NT}^1 & \dots & x_{NT}^k \end{pmatrix}_{NT\times K},$$

$$\varepsilon = \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1,T} \\ \vdots \\ \varepsilon_{N,T} \end{pmatrix}_{NT\times 1}, \quad a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix}_{N\times 1}, \quad \underline{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}_{N\times 1}, \quad D = I_N \otimes l_T, ,$$

where l_T is a $(T \times 1)$ unit vector.

For convenience, we define two matrixes used often in this chapter:

$$W_n = I_N \otimes (I_T - \frac{J_T}{T}), \text{ and } \overline{B}_n = I_N \otimes \frac{J_T}{T},$$

where I_T is T-order identity matrix and $J_T = (1, \dots, 1)'_{1\times T}$.

The plan of this chapter is as follows. Section 2 considers the inconsistency of LSDV estimator when T is finite. Section 3 considers several estimators from the setup of dynamic error component models when the unobserved effects is assumed to be random. Section 4 consider the MLE estimator in consideration of the initial conditions. Section 5 show the gain of efficiency from the GMM estimator by imposing the extra moment conditions and the restrictions on the initial conditions. Section 6 gives some concluding summary.

1.2 The Inconsistency of the LSDV Estimator

In the static case in which all the explanatory variables are exogenous and are uncorrelated with the effects, the OLS estimator, although possibly less efficient, is still unbiased and consistent. But in the dynamic case the correlation between the lagged dependent variable and individual-specific effects would seriously contaminate the property of OLS estimator. We will show the bias of the least-squares dummy-variables (LSDV) estimator for a dynamic fixed-effects model and then see how to treat with the problem. We assume that the disturbances satisfy the conditions as follows:

$$\begin{aligned} E(\varepsilon_{it}|y_{i,t-1},\ldots,y_{i0},X_i) &= 0 \\ V(\varepsilon_{it}|y_{i,t-1},\ldots,y_{i0},X_i) &= \sigma_\varepsilon^2 \text{ for all } i \text{ and } t \end{aligned} \tag{1.3}$$

i.e., the disturbances have a zero conditional mean (which implies they are serially uncorrelated), and are homoscedastic. In the traditional fixed effects approach, a_i is treated as a scalar parameter to be estimated. Multiplying the equation (1.2) by W_n,

$$W_nY = \rho W_nY_{-1} + W_nX\beta + W_n\varepsilon, \tag{1.4}$$

Because W_n is a symmetric idempotent matrix, the LSDV estimators for ρ, β can be expressed as the within estimator are as follows:

$$\begin{pmatrix} \hat{\rho} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} Y_{-1}'W_nY_{-1} & Y_{-1}'W_nX \\ X'W_nY_{-1} & X'W_nX \end{pmatrix}^{-1} \begin{pmatrix} Y_{-1}'W_nY_{-1} \\ X'W_nY_{-1} \end{pmatrix}. \tag{1.5}$$

The estimator of unobserved effects is as follows:

$$\hat{a}_i = \bar{y}_i - \hat{\rho}\bar{y}_{i,-1} - \bar{x}_i\hat{\beta}, \quad i = 1,\ldots,N, \tag{1.6}$$

where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$ and $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$.

When $N \to \infty$ with fixed T, given the above assumption of (1.3) on the disturbance, we write the equation of (1.5) in the probability limit as follows:

$$\begin{aligned} \operatorname*{plim}_{N\to\infty} \begin{pmatrix} \hat{\rho} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} \rho \\ \beta \end{pmatrix} &+ \begin{pmatrix} \operatorname*{plim}_{N\to\infty} \frac{1}{NT}Y_{-1}'W_nY_{-1} & \operatorname*{plim}_{N\to\infty} \frac{1}{NT}y_{-1}'W_nX \\ \operatorname*{plim}_{N\to\infty} \frac{1}{NT}X'W_nY_{-1} & \operatorname*{plim}_{N\to\infty} \frac{1}{NT}X'W_nX \end{pmatrix}^{-1} \\ &\times \begin{pmatrix} \operatorname*{plim}_{N\to\infty} \frac{1}{NT}Y_{-1}'W_n\varepsilon \\ \operatorname*{plim}_{N\to\infty} \frac{1}{NT}X'W_n\varepsilon \end{pmatrix} \end{aligned} \tag{1.7}$$

We prove $\operatorname*{plim}_{N\to\infty} \frac{1}{NT} Y'_{-1} W_n \varepsilon \neq 0$ in the following. The inconsistency of this estimator rely on the fact that, given the assumption about the disturbances, one has $\operatorname*{plim}_{N\to\infty} \frac{1}{NT} X' W_n \varepsilon = 0$ under the strict exogeneity assumption, but

$$\begin{aligned}
\operatorname*{plim}_{N\to\infty} \tfrac{1}{NT} Y'_{-1} W_n \varepsilon &= \operatorname*{plim}_{N\to\infty} \tfrac{1}{NT} \textstyle\sum_{i=1}^{N} \sum_{t=1}^{T} (y_{i,t-1} - \bar{y}_{i,-1})(\varepsilon_{it} - \bar{\varepsilon}_i) \\
&= -\operatorname*{plim}_{N\to\infty} \tfrac{1}{N} \textstyle\sum_{i=1}^{N} \bar{y}_{i,-1} \bar{\varepsilon}_i \\
&= -\frac{\sigma_a^2}{T^2} \frac{(T-1) - T\rho + \rho^T}{(1-\rho)^2} \neq 0
\end{aligned} \tag{1.8}$$

Equation (1.7) can be rewritten as follows:

$$\begin{aligned}
\operatorname*{plim}_{N\to\infty} \begin{pmatrix} \hat{\rho} - \rho \\ \hat{\beta} - \beta \end{pmatrix} &= \Delta \times \begin{pmatrix} \operatorname*{plim}_{N\to\infty} \frac{1}{NT} Y'_{-1} W_n \varepsilon \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} \Delta_{11} \cdot \operatorname*{plim}_{N\to\infty} \frac{1}{NT} Y'_{-1} W_n \varepsilon \\ \Delta_{21} \cdot \operatorname*{plim}_{N\to\infty} \frac{1}{NT} Y'_{-1} W_n \varepsilon \end{pmatrix},
\end{aligned} \tag{1.9}$$

where $\Delta = \begin{pmatrix} \operatorname*{plim}_{N\to\infty} \frac{1}{NT} Y'_{-1} W_n Y_{-1} & \operatorname*{plim}_{N\to\infty} \frac{1}{NT} y'_{-1} W_n X \\ \operatorname*{plim}_{N\to\infty} \frac{1}{NT} X' W_n Y_{-1} & \operatorname*{plim}_{N\to\infty} \frac{1}{NT} X' W_n X \end{pmatrix}^{-1}$.

When T is kept fixed, the LSDV estimator of an AR(1) fixed effects model is not consistent. The inconsistency mainly comes from the fact that correlation exists between $(y_{i,t-1} - \bar{y}_i)$ and $(\varepsilon_{it} - \bar{\varepsilon}_i)$. In other words, the individual means, $\overline{y}_i$ and ε_{it} are correlated with each other, although the past of $y_{i,t}$ and ε_{it} are uncorrelated. As it is clear from equation (1.8), when N and $T \to \infty$, this estimator is consistent since $\operatorname*{plim}_{N\to\infty} \frac{1}{NT} Y'_{-1} W_n \varepsilon = 0$. Unfortunately, most panel data sets of interest contain small number of time-periods. Therefore, we should look for estimation methods that are consistent when T is fixed. A traditional way to tackle the problem of within estimator is to use an instrumental variables estimation method after a transformation to estimate a_i. To be more precise, using an appropriate transformation and then IV can implemented to consistently estimate the parameters.

Balestra and Nerlove (1966) have shown that Two-Stage Least Squares which uses current and lagged values of x_{it} as instrument variables is available. Based on the model of (1.2), let us define the complete set of instruments as $Z^* = (D, Z)$, where D can is the set of dummy variables accounting for the individual effects. An appropriate transformation for (1.2) is as follows:

$$P_{Z^*}Y = \rho P_{Z^*}Y_{-1} + P_{Z^*}X\underline{\beta} + \ P_{Z^*}D\underline{a} + P_{Z^*}\varepsilon, \tag{1.10}$$

where $P_{Z^*} = Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}$, which is projector onto the space spanned by Z^*. By the Frisch-Waugh theorem, the solution to the problem amounts to applying the OLS to the equation as follows:

$$W_nP_{Z^*}Y = \rho W_nP_{Z^*}Y_{-1} + W_nP_{Z^*}X\underline{\beta} + W_nP_{Z^*}\varepsilon, \tag{1.11}$$

where $W_n = I - P_D = I - D(D'D)^{-1}D' = I_N \otimes (I_T - \frac{J_T}{T})$. The fixed effects a_i can be "estimated" as $\tilde{a} = P_D(Y - Y_{-1}\rho - X\underline{\hat{\beta}})$. If we add the assumption on the error terms as follows:

$$\varepsilon_{it} \text{ are independently and identically distributed with mean 0 and variance } \sigma_\varepsilon^2, \tag{1.12}$$

the property of $\sqrt{N}$-asymptotic normality is valid, i.e.,

$$\sqrt{N}(\underline{\hat{\theta}} - \underline{\theta}) \sim \text{Normal}(0, \sigma_\varepsilon^2(\operatorname*{plim}_{N\to\infty} \frac{1}{\mathrm{N}}\mathrm{\tilde{X}'W_nZ(Z'W_nZ)^{-1}Z'W_n\tilde{X}})^{-1}), \tag{1.13}$$

where $\underline{\theta} = (\rho, \underline{\beta})$ and $\tilde{X} = (Y_{-1}, X)$.

Anderson-Hsiao (1982) have proposed to use as instrument variables the lagged first-difference of dependent variable or the level of dependent variable lagged two or more periods after first-difference transformation into equation (1.2) as follows:

$$\Delta Y = \rho\Delta Y_{-1} + \Delta X\underline{\beta} + \Delta\varepsilon, \tag{1.14}$$

It is obvious that the variable $y_{i,t-2}$ (or lagged more periods) and $\Delta y_{i,t-2}$ are valid instruments since they are correlated with $\Delta y_{i,t-1}$ but uncorrelated with the disturbance $\Delta\varepsilon_{it}$.

Arellano (1988) considers a specific model allowing for only one exogenous variable which follows a stationary AR(1) process plus a lagged endogenous variable and he has shown that the variance of the estimator using $\Delta y_{i,t-2}$ as instrument variable can be very high due to near-singular matrices entering its definition. Arellano (1988) proposed $y_{i,t-2}$ instead of $\Delta y_{i,t-2}$ as the instrumental variable. Given the assumptions (1.3) and (1.12), the property of $\sqrt{N}$-asymptotic normality is valid. We write the asymptotic distribution as follows:

$$\sqrt{N}(\hat{\underline{\theta}} - \underline{\theta}) \sim \text{Normal}\left(0, \sigma_\varepsilon^2 (\operatorname*{plim}_{N\to\infty} \frac{1}{\text{N}}((\tilde{\text{Z}}\Delta\tilde{\text{X}}')^{-1}\tilde{\text{Z}}\Psi\tilde{\text{Z}}(\Delta\tilde{\text{X}}'\tilde{\text{Z}})^{-1}))\right), \qquad (1.15)$$

where $Z = [Z_1, \ldots, Z_N]'$, $\tilde{X} = [Y_{-1}, X]$

and

$$\Psi = I_N \otimes \Sigma_D = I_N \otimes \begin{pmatrix} 2 & -1 & & 0 \\ -1 & 2 & & \\ & \ddots & \ddots & \\ & & & 1 \\ 0 & & -1 & 2 \end{pmatrix}, \qquad (1.16)$$

since the disturbance in model (1.14) MA(1).

There exists estimators more efficient than that of (1.15) since the disturbance in model (1.14) is MA(1). Transforming the equation (1.14) by multiplying $\Psi^{\frac{-1}{2}}$, we have an equation as follows:

$$\Psi^{\frac{-1}{2}}\Delta Y = \rho\Psi^{\frac{-1}{2}}\Delta Y_{-1} + \Psi^{\frac{-1}{2}}\Delta X\underline{\beta} + \Psi^{\frac{-1}{2}}\Delta\varepsilon. \qquad (1.17)$$

Sevestre (1992) suggested using as instruments y_{it-2} or $\Delta y_{i,t-2}$, plus the current and lagged values of ΔX or $\Psi^{\frac{-1}{2}}\Delta X$, provided that X is strictly exogenous. Such an estimator is more efficient than the one using the same instruments on the untransformed equation (1.14) (see White [1984]). Given assumptions (1.3) and (1.12), the property of $\sqrt{N}$-asymptotic normality is as follows:

$$\sqrt{N}(\hat{\underline{\theta}} - \underline{\theta}) \sim \text{Normal}\left(0, \sigma_\varepsilon^2[\underset{N\to\infty}{\text{plim}} \frac{1}{\text{N}}((\Delta\tilde{\text{X}}\Psi^{-1}\tilde{\text{Z}})^{-1}\tilde{\text{Z}}'\Psi^{-1}\tilde{\text{Z}}(\tilde{\text{Z}}'\Psi^{-1}\Delta\tilde{\text{X}})^{-1})]\right). \quad (1.18)$$

Nevertheless, $\Psi^{\frac{-1}{2}}\Delta\varepsilon$ means that the disturbances are linear combination of ε_{it} and hence the lagged values of y_{it-2} or $\Delta y_{i,t-2}$ are no longer valid instruments except for y_{i0}, but nothing can be said about the relative performance of this estimator and the ones suggested by Anderson-Hsiao (1982), since the instrument is different.

Based on the differenced equation (1.14), Arellano-Bond (1991) proposed another way to find a more efficient estimator, generalized instrumental variables estimator, which contains all the orthogonality conditions that exist between lagged values of the endogenous variables and the disturbances. Assumption (1.3) implies

$$E(y_{is}\Delta\varepsilon_{it}) = 0, t = 2, \ldots, T, s = 0, \ldots, t-2. \quad (1.19)$$

At period of t, $y_{i0}, y_{i1}, \ldots, y_{i,t-2}$ are valid instruments for $\Delta y_{i,t-1}$, respectively. Because X is assumed to be strictly exogenous variables, Δx_{it} is a valid instrument for itself. Then the complete set of instrument variables can be defined as

$$\tilde{Z}_i = \begin{pmatrix} y_{i0} & 0 & 0 \ldots & 0 & \Delta x_{i2} & 0 & \ldots & 0 \\ 0 & (y_{i0}, y_{i1}) & 0 \ldots & 0 & 0 & \Delta x_{i3} & & 0 \\ \vdots & & (y_{i0}, y_{i1}, y_{i2}) & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & \ldots & (y_{i0}, y_{i1}, \ldots, y_{i,T-2}) & 0 & \ldots & \ldots & \Delta x_{iT} \end{pmatrix} \quad (1.20)$$

The generalized IV estimator is defined as

$$\hat{\underline{\theta}} = \left(\Delta\tilde{X}' P_{\tilde{Z}} \Delta\tilde{X}\right)^{-1} \left(\Delta\tilde{X}' P_{\tilde{Z}} \Delta y_{-1}\right), \quad (1.21)$$

where

$$P_{\tilde{Z}} = \tilde{Z}\Gamma\tilde{Z}' \; with \; \Gamma = (\tilde{Z}\Psi\tilde{Z}')^{-1} = (\frac{1}{N}\sum_{i=1}^{N}\tilde{Z}_i'\Sigma_D\tilde{Z}_i)^{-1}$$

Actually, the estimator is to apply GLS to the model (1.13) multiplied by $\tilde{Z}'$ as follows:

$$\tilde{Z}'\Delta Y = \tilde{Z}'\Delta Y_{-1}\rho + \tilde{Z}'\Delta X\underline{\beta} + \tilde{Z}'\Delta\varepsilon. \tag{1.22}$$

Since the ε_{it} is not autocorrelated in this model, the estimator is the most efficient within the class of instrumental variables estimators using lagged value of y_{it} as instruments. Given assumption (1.3) and (1.12), Its property of $\sqrt{N}$-asymptotic normality is as follows:

$$\sqrt{N}(\hat{\underline{\theta}} - \underline{\theta}) \sim \text{Normal}\left(0\ ,\ \ \operatorname*{plim}_{N\to\infty} \frac{1}{\mathrm{N}}\left(\Delta\tilde{\mathrm{X}}'\mathrm{P}_{\tilde{\mathrm{Z}}}\Delta\tilde{\mathrm{X}}\right)^{-1}\right). \tag{1.23}$$

We can write GMM estimator by replacing $P_{\tilde{Z}}$ with $P^*_{\tilde{Z}}$, where $P^*_{\tilde{Z}} = \tilde{Z}\Gamma\tilde{Z}'$ with $\Gamma = (\frac{1}{N}\sum_{i=1}^{N}\tilde{Z}_i\underline{\nu}_i\underline{\nu}_i'\tilde{Z}_i)^{-1}$ where $\underline{\nu}_i$ is the vector of disturbances of the differenced equation (1.14). Nevertheless, to ensure that the instrument of (1.20) is valid, the order of autocorrelation of disturbances is required to be not greater than one.

The choice of these various estimators in estimating an AR(1) fixed model depends on two main criteria: the degree of serial correlation of the ε_{it} disturbance terms and the exogeneity of X. We end the section with some conclusion as follows:

1. When the values of ε_{it} are correlated and x_{it} is strictly exogenous the Two-Stage Least Squares estimator which use the current and lagged values of the x_{it} as IV (Balestra-Nerlove, 1966) is better one.

2. If the values of ε_{it} are correlated while x_{it} is still strictly exogenous then the generalized IV estimator proposed by Arellano-Bond (1991) is better than others.

3. If the values of ε_{it} are correlated and x_{it} is not strictly exogenous then using lagged values of the Δx_{it} as IV for estimating the model (1.14) is preferred.

1.3 Estimators of error components model

The section considers estimation of the AR(1) model under the assumption that the unobserved effects are always random. The model of (1.1) or (1.2) is adopted in this section and we use the following assumptions:

$$\begin{aligned} &E(a_i) = E(\varepsilon_{it}) = 0, \text{ for all } i \text{ and } t, \\ &E(a_i x_{it}) = 0, \text{ for all } i \text{ and } t, \\ &E(a_i \varepsilon_{it}) = 0, \text{ for all } i \text{ and t}, \\ &E(a_i a_j) = \begin{cases} \sigma_a^2 & i = j, \\ 0 & i \neq j. \end{cases} \\ &E(\varepsilon_i \varepsilon_i') = \sigma_\varepsilon^2 I_T, \text{ where } \varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})', \text{ for all } i. \end{aligned} \tag{1.24}$$

(1.24) implies that the special second-order structure of disturbances in model (1.2), $u_{it} = a_i + \varepsilon_{it}$ or $u_i = a + \varepsilon_i$, as follows:

$$Var(u_i) = E(u_i u_i') = \Omega_T = \sigma_\varepsilon^2 W_n + (\sigma_\varepsilon^2 + T\sigma_a^2)\overline{B}_n = \sigma_\varepsilon^2 (W_n + \frac{1}{\theta}\overline{B}_n),$$

with $\theta^2 = \sigma_\varepsilon^2/(\sigma_\varepsilon^2 + T\sigma_a^2)$. Under the specific assumption we can obtain the GLS estimator by simply imposing OLS on the equation (1.2) multiplied $(W_n + \theta\overline{B}_n)$:

$$(W_n + \theta\overline{B}_n)Y = \rho(W_n + \theta\overline{B}_n)Y_{-1} + (W_n + \theta\overline{B}_n)X\underline{\beta} + (W_n + \theta\overline{B}_n)U, \tag{1.25}$$

where $U = (u_1, \dots, u_N)'$. Nevertheless, the GLS estimator is not the most efficient estimator because it does not impose any restrictions on the relation between y_{i0} and a_i or ε_{it}. To shed light on the importance of the initial value, we write substitution recursively in equation from (1.1) to obtain:

$$y_{it} = \rho^t y_{i0} + \sum_{j=1}^{t} \rho^{j-1} x_{i,t-j+1}\beta_i + \frac{1-\rho^t}{1-\rho} a_i + \sum_{j=1}^{t} \rho^{j-1}\varepsilon_{i,t-j+1}. \tag{1.26}$$

Each observation on the endogenous variable can be expressed as a linear combination of four variables: $\rho^t y_{i0}$, $\sum_{j=1}^t \rho^{j-1} x_{i,t-j+1}\beta_i$, $\frac{1-\rho^t}{1-\rho} a_i$, and $\sum_{j=1}^t \rho^{j-1}\varepsilon_{i,t-j+1}$. The first term $\rho^t y_{i0}$ depends on the initial value. At this stage, it is clear that the initial values do influence the asymptotic behavior of estimator as long as T is finite and ρ is not zero. Theoretically, the first date of the sample is arbitrarily chosen and we cannot easily justify a different treatment of the first and the subsequent observations. For example, we assume $y_{it} = f(a_i, \varepsilon_{it}, \varepsilon_{i,t-1}, \ldots)$. This means that the outcome on y in time t depends on the individual effects a_i and on a serially uncorrelated disturbance ε_{it}. Therefore, if unobserved effects are non-random, then the initial observations are also non-random; and on the contrary, if the unobserved effects are random, then the initial observations are random.

As a practical matter, the assumption that the unobserved effects are non-random means that the initial observations are independent of the exogenous variables and the unobserved effects, usually an untenable assumption. The interpretation of the relation between the initial observation and the unobserved effects characterize the dynamic panel data with random-effect formulation. We assume that y_{i0} is identically and independently distributed variables characterized by the second order moment $E(y_{i0}^2)$ and the correlation with a_i, $E(y_{i0}a_i)$. Replacing θ with $\sqrt{\lambda}$, Maddala (1971) proposed λ-class estimator and have shown that all usual error-component estimators belong to such an estimator for a AR(1) framework. For example, the within estimator has $\lambda = 0$; OLS if $\lambda = 1$; GLS estimator if $\lambda = \theta^2$; If ρ =0, then, obviously, all these estimators are consistent, while almost all λ-class estimators are not consistent if $\rho \neq 0$.

Under the above assumption on the distribution of y_{i0} the asymptotic bias of any λ-class estimator is dependent on $E(y_{i0}^2)$ and $E(y_{i0}a_i)$. This shows that assumptions on the initial observations do influence the magnitude of the bias of these estimators. The main result is as follows: whatever $E(y_{i0}^2)$ and $E(y_{i0}a_i)$ are, $\operatorname*{plim}_{N\to\infty} \rho(\lambda)$ is an increasing function of λ and hence the relation is assured as follows:

$$\operatorname*{plim}_{N\to\infty} \hat{\rho}(0) < \rho < \operatorname*{plim}_{N\to\infty} \hat{\rho}(\theta^2) < \operatorname*{plim}_{N\to\infty} \hat{\rho}(1) < \operatorname*{plim}_{N\to\infty} \hat{\rho}(\infty). \tag{1.27}$$

It is obvious that there exists a value $\lambda^* \in [0, \theta^2]$ such that $\operatorname*{plim}_{N\to\infty} \hat{\rho}(\lambda^*) = \rho$. Sevestre-Trognon (1983) have given the value of λ^* as follows:

$$\lambda^* = K(1-\delta)/(\frac{1-\rho^T}{1-\rho}\frac{E(y_{i0}a_i)}{\sigma_a^2+\sigma_\varepsilon^2} + K(1-\delta+T\delta)), \tag{1.28}$$

with $K = (T-1-T\rho+\rho^T)/T(1-\rho)^2$, $\delta = \sigma_a^2/\sigma_a^2+\sigma_\varepsilon^2$.

When $\mathrm{E}(y_{i0}a_i) = 0$, λ^* is equal to θ^2, which means that the consistent estimator λ-class estimator is the GLS estimator. Usually, $\lambda^* \neq \theta^2$, which confirms GLS is not consistent in such a model.

It is worth noticing that sometimes the λ-class estimator cannot be thought of as an estimator because of unknown parameter λ, which leads to two-stage estimation; and hence the property of $\sqrt{N}$-asymptotic normality heavily depends on the asymptotic property of $\hat{\lambda}$. $\hat{\rho}(\lambda^*)$ and $\hat{\beta}(\lambda^*)$ with λ^* defined as in (1.28)are derived from the AR(1) model and such an estimator cannot be extended to AR(p) models.

GMM procedures are thus possible to impose some restrictions to find more efficient estimators. We assume $\beta_i = \beta$ for all i. By adding the assumption that

$y_{i0} = C + \alpha_1 a_i + \alpha_2 \varepsilon_{i0}$

proposed by Anderson-Hsiao (1982) into the system (1.2), the system is a triangular model with $T+1$ endogenous variables $y_{i0}, \ldots, y_{iT}$ and $T+1$ exogenous variables (C and $x_{i1}, \ldots, x_{iT}$). We can express the $T+1$ equations by compact matrix form as follows:

$$A\underline{y}_i - B\underline{x}_i = \underline{\eta}_i, \tag{1.29}$$

where

$$\underline{y}_i = (y_{i0}, y_{i1}, \ldots, y_{iT})', \ \underline{x}_i = (1, x_{i1}, \ldots, x_{iT})',$$

$$\underline{\eta}_i = (\alpha_1 a_i + \alpha_2 \varepsilon_{i0}, a_i + \varepsilon_{i1}, \ldots, a_i + \varepsilon_{iT})',$$

and

$$A = \begin{pmatrix} 1 & 0 & \dots & \dots & 0 \\ -\rho & 1 & 0 & \dots & 0 \\ 0 & -\rho & 1 & \dots & 0 \\ \vdots & \vdots\ddots & \vdots\ddots & \vdots\ddots & \vdots \\ 0 & \dots & 0 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} C & 0 & \dots & 0 \\ 0 & \beta & \dots & \vdots \\ \vdots & \vdots\ddots & \vdots\ddots & 0 \\ 0 & \dots & \dots & \beta \end{pmatrix}$$

where A and B are $(T+1)\times(T+1)$ matrix. The structure of disturbances are defined as $E(\eta_i) = 0$ and $Var(\underline{\eta}_i) = \Omega$ and hence

$$\Omega = \begin{pmatrix} \omega & \tau' \\ \tau & V \end{pmatrix}, \quad V = \sigma_a^2 J_T + \sigma_\varepsilon^2 I_T,$$

$$\omega = \alpha_1^2\sigma_a^2 + \alpha_2^2\sigma_\varepsilon^2, \quad \tau' = \alpha_1\sigma_a^2(1,\dots,1).$$

In the simplified case, we find that the IV estimator is consistent. If the disturbance have an error component structure, the 3SLS estimator is not as efficient as full information maximum likelihood (FIML) estimator. If the variance-covariance matrix Ω is unconstrained, then 3SLS and FIML are fully efficient. Based on the differenced model, several IV or GMM suggest consistent and efficient estimator, but their relative efficiencies are hard to determine. For example, the GMM estimator with asymptotic efficiency may not often perform better than the Balestra-Nerlove estimator in finite samples (See Sevestre-Trognon [1990]). We found the interpretation of the initial conditions make it possible to obtain more efficient GMM estimators. The subsequent section we discuss the MLE estimators with the different treatments of y_{i0}. And then we introduce the initial conditions into GMM to obtain more efficient estimators.

1.4 Properties of the ML estimator

According to (1.10), the two-stage estimators of Balestra and Nerlove (1966) is OLS applied to equation (1.11). Such an estimator is equal to MLE assuming that y_{i0} is non-random. Because this assumption implies that the initial observation is discarded from the system, the ML estimator is not the unconditional ML estimator. The asymptotic correlation of y_{i0} and a_i contaminate the consistency of Balestra-Nerlove estimator when the nonrandom assumption of the initial observation is dropped. A natural solution is to construct the full likelihood function which includes the initial observation to obtain the unconditional ML estimator. Barghava-Sargan (1983) proposed an unconditional estimator by considering the framework (1.1) into which the observed individual variables z_i will be introduced. We write the model as follows:

$$\begin{aligned} y_{it} = \rho\, y_{i,t-1} + x_{it}\beta + z_i\gamma + u_{it}, \quad & u_{it} = a_i + \varepsilon_{it}, \\ & i = 1, ..., N \;;\; t = 1, ..., T. \end{aligned} \tag{1.30}$$

The initial values are assumed to follow:

$$y_{i0} = z_i\phi + \nu_{i0}. \tag{1.31}$$

Such a formulation has been adopted by Chamberlain (1984) and Blundell and Smith (1991) among others. The unobserved effects is assumed to be as follows:

$$a_i = \psi\nu_{i0} + c_i, \tag{1.32}$$

where c_i is independent of ν_{i0}. In this model $(\nu_{i0}, c_i, \varepsilon_{i1}, \ldots, \varepsilon_{iT})$ are distributed as $\text{Normal}(0, \text{diag}(\sigma^2_{\text{y0}}, \sigma^2_{\text{a}}, \sigma^2_{\varepsilon} 1'_{\text{T}}))$ and the log-likelihood function is:

$$\begin{aligned} \mathcal{L}(\rho, \beta, \gamma, \psi, \sigma^2_{y0}, \sigma^2_a, \sigma^2_\varepsilon) = & -\frac{NT}{2}\log 2\pi - \frac{N}{2}\log|\Omega| - \frac{N}{2}\log\sigma^2_a \\ & -\frac{1}{2}\sum_{i=1}^N \tau_i'\Omega^{-1}\tau_i - \frac{1}{2\sigma^2_{y0}}\sum_{i=1}^N \nu^2_{i0}, \end{aligned} \tag{1.33}$$

with

$\tau_i' = (y_{i1} - \rho y_{i0} - x_{i1}\beta - z_i\gamma - \psi\nu_{i0}, \ldots, y_{iT} - \rho y_{i,T-1} - x_{iT}\beta - z_i\gamma - \psi\nu_{i0})$

$\nu_{i0} = y_{i0} - z_i\phi,$

$\Omega = \sigma_\varepsilon^2 W_n + (\sigma_\varepsilon^2 + T\sigma_a^2)\overline{B}_n.$

The ML estimators solve the normal equations:

$$\partial\mathcal{L}(\rho, \beta, \gamma, \psi, \sigma_{y0}^2, \sigma_a^2, \sigma_\varepsilon^2)/\partial(\rho, \beta, \gamma, \psi, \sigma_{y0}^2, \sigma_a^2, \sigma_\varepsilon^2) = 0. \tag{1.34}$$

The ML estimators of ϕ and σ_{y0}^2 are solved by the equation from (1.34) as follows:

$$\partial\mathcal{L}/\partial\phi = \frac{1}{(\sigma_\varepsilon^2 + T\sigma_a^2)} \sum_{i=1}^{N} u_{i0} l_T' \tau_i = 0$$

and

$$\partial\mathcal{L}/\partial\sigma_{y0}^2 = -\frac{N}{\sigma_{y0}^2} + \frac{1}{2\sigma_{y0}^4} \sum_{i=1}^{N} u_{i0}^2 = 0.$$

These imply that these two estimators are OLS estimators on the equation (1.31). The other estimators can be solved by replaced the residual $\hat{u}_{i0}$ in τ_i on equation (1.34). This approach turn out to be two-stage estimation from which we can solve the ML estimation of (1.30) where the unobserved u_{i0} is replaced by the residual $\hat{u}_{i0}$ obtained from the OLS estimation on (1.31)in advance. Obviously, if we add another term $x_{i0}\alpha$ into the equation (1.31), the initial observations are defined as follows:

$$y_{i0} = \phi z_i + x_{i0}\alpha + \nu_{i0} \tag{1.35}$$

and (1-30) are unchanged, i.e., the variable x_{i0} does not enter the autoregressive equation and hence the simple split ion between OLS and MLE disappears. Sevestre-Trognon (1990) proposed an auxiliary model:

$$y_{it} = \rho\, y_{i,t-1} + x_{it}\beta + z_i\gamma + x_{i0}\delta + u_{it}. \tag{1.36}$$

The ML estimation of ρ, β, γ, ϕ, ψ, δ, σ_{y0}^2, σ_a^2, and σ_ε^2 are asymptotically equivalent to the ML estimators of (1.30). If $\hat{\alpha}$ is OLS estimator of α on (1.35), it is shown that $\hat{\alpha}^* =$

$\hat{\alpha} + \frac{\hat{\delta}^*}{\hat{\psi}^*}$ is asymptotically efficient if $\hat{\delta}^*$ and $\hat{\psi}^*$ are the ML estimators of δ and ψ in the auxiliary equation (1.36) (See Sevestre-Trognon [1990], and Blundell-Smith [1991]). Such a ML estimation suggests the asymptotically most efficient estimators when the disturbances are normal. The consistency of MLE depends on the assumptions on the initial observations. The other less restrictive assumption on y_{i0} is to leave the correlation of initial observations and unobserved effects $E(y_{i0}a_i)$ free and consider it as a parameter to be estimated. By specifying a distribution of y_{i0} with mean μ_{y0} and variance σ^2_{y0}, as well as $cov(y_{i0}, a_i) = \underline{\phi}\sigma^2_{y0}$, Anderson-Hsiao (1982) solved out the unconditional ML estimators from the model (1.30) and studied the consistency properties of the MLEs for dynamic model with a random-effect formulation. We give a summary in Table(1.1). It is obvious that the properties of ML estimators for dynamic random-effects models depends on the assumptions on the initial conditions; so do those of GLS and GMM.

1.5 The efficiency of GMM estimator

The previous IV methods for estimating the dynamic panel data model (e.g. Anderson and Hsiao [1981]; Hsiao [1986]; Arellano [1988]; Arellano and Bond [1991]) first-difference the equation to remove the unobserved effects, and then use instrumental variables, using as instruments values of the dependent variables lagged two or more periods. More recent papers (Ahn and Schmidt [1995]; Arellano and Bover [1995]; Blundell and Bond [1998]; Hahn [1999]) proposed additional moment conditions, often not exploited by these estimators, that can be used in generalized method of moments (GMM) estimation to improve upon these IV estimators.

For expository purpose, we consider the simple dynamic panel data model, which does not contain any additional regressor beyond the lagged dependent variable, will be used often to express the available moment conditions.[1] We write the simple model as follows:

$$y_{it} = \rho y_{i,t-1} + u_{it}, \quad u_{it} = a_i + \varepsilon_{it}. \tag{1.37}$$

Ahn and Schmidt (1997) assume that all variables across individual are independent. Various subsets of assumption about initial conditions and the errors are made in the following:

(i) For all i, ε_{it} is uncorrelated with y_{i0} for all t.

(ii) For all i, ε_{it} is uncorrelated with c_i for all t.

(iii) For all i, ε_{it} are mutually uncorrelated.

(iv) For all i, $\text{var}(\varepsilon_{it})$ is the same for all t.

Given the four assumptions (i)- (iv), there are several plausible cases of 16 possible combinations corresponding to imposing or not imposing each of then. We just discuss the case imposed by assumption (i)-(iv) to explain the exploitation of additional moment conditions. Let Σ be the covariance matrix of $(\varepsilon_{i1}, \ldots, \varepsilon_{iT}, y_{i0}, a_i)$ (see Ahn and Schmidt [1995b] Eq.(6)). Let Λ be the covariance matrix of $(u_{i1}, \ldots, u_{iT}, y_{i0})$ (see Ahn and Schmidt [1995b] Eq.(7)). Assumptions (i) - (iv) imply that we have $[(T-1)+(T-1)+(T(T-1)/2-1)]$ restrictions on Σ as follows:

$$\begin{aligned}
\sigma_{\alpha t} &\equiv E(\varepsilon_{it} a_i) = 0, \text{ for all } t, \\
\sigma_{0t} &\equiv E(y_{i0}\varepsilon_{it}) = 0, \text{ for all } t, \\
\sigma_{ts} &\equiv E(\varepsilon_{it}\varepsilon_{is}) = 0, \ t \neq s, \\
\sigma_{tt} &\equiv E(\varepsilon_{it}\varepsilon_{it}) = 0, \text{ for all } t.
\end{aligned} \tag{1.38}$$

[1] The moment conditions implied by exogeneity assumptions on additional regressors have been identified by Shmidt, Ahn, and Wyhowski (1992), Ahn and Schmidt (1995b), and Arrelano and Bover (1995)

(1.38) implies restrictions on Λ in three types:

$$\lambda_{0t} \equiv E(y_{i0}u_{it}), \ t = 1, \dots, T,$$
$$\lambda_{ts} \equiv E(u_{it}u_{is}), \ t \neq s; \ t, s = 1, \dots, T, \tag{1.39}$$
$$\lambda_{tt} \equiv E(u_{it}^2) = \sigma_u^2, \ t = 1, \dots, T.$$

These restrictions corresponding to the moment conditions are as follows:

$$E(y_{i0}\Delta u_{it}) = 0, \qquad \text{for } t = 2, \dots, T,$$
$$E(u_{i1}^2 - u_{it}^2) = 0, \qquad \text{for } t = 2, \dots, T, \tag{1.40}$$
$$E(u_{it}u_{is} - u_{i2}u_{i1}), \qquad \text{for } t = 3, \dots, T, s < t.$$

The conditions of (1.40) are algebraically equal to the following:

$$E(y_{is}\Delta u_{it}) = 0, \quad \text{for } t = 2, \dots, T, \ s = 0, \dots, t-2,$$
$$E(u_{iT}\Delta u_{it}) = 0, \quad \text{for } t = 2, \dots, T-1, \tag{1.41}$$
$$E(\overline{u}_i\Delta u_{it}) = 0, \quad \text{for } t = 2, \dots, T,$$

where $\overline{u}_i = T^{-1}\sum_{t=1}^{T} u_{it}$. The first condition of (1.41) is derived from the differenced equation of

$$\Delta y_{it} = \rho\Delta y_{i,t-1} + \Delta\varepsilon_{it} \quad \text{for} t \ = \ 2, \dots, T \tag{1.42}$$

Holtz-Eakin (1988), Holtz-Eakin, Newey, and Rosen (1988), and Arellano and Bond (1991) have shown that the valid instrument variables for equation (1.42) is $(y_{i0}, \dots, y_{i,t-2})$ which implied by the first condition of (1.41). Given assumptions (i) - (iv), Ahn and Schmidt (1995b) show that the second condition of (1.41) can be replaced by

$$E(y_{i,t-2}\Delta u_{i,t-1} - y_{i,t-1}\Delta u_{it}) = 0, \ \ t = 3, \dots, T, \tag{1.43}$$

which are linear in ρ.

To derive the general result, Ahn and Schmidt consider the model including exogenous variables. We write its compact matrix form as

$$Y = Y_{-1}\rho + X\ \beta\ + Z\ \gamma\ +\ U = W\xi + U, \tag{1.44}$$

where $U = a + \varepsilon$, X is time-varying explanatory variables, Z time-invariant explanatory variables and Y_{-1} lagged dependent variables. We write the T observation for individual i as $u_i(\xi) = y_i - W_i\xi$ to emphasize the dependence of u_i on ξ. Exogeneity assumptions on x_i and z_i generate linear moment conditions of the form

$$E[R_i' u_i(\xi)] = 0, \tag{1.45}$$

where R_i is a function of the exogenous variables and ξ is $(\rho, \beta', \gamma')'$, $E(y_{is}\Delta u_{it})$ leads to the moment conditions being a linear function of ξ and can be written as $E(A_i' u_i(\xi))$=0, where A_i is the $T \times T(T-1)/2$ matrix

$$A_i = \begin{bmatrix} -y_{i0} & 0 & 0 & \dots & 0 \\ y_{i0} & -(y_{i0}, y_{i1}) & 0 & \dots & 0 \\ 0 & (y_{i0}, y_{i1}) & -(y_{i0}, y_{i1}, y_{i2}) & \dots & 0 \\ 0 & 0 & (y_{i0}, y_{i1}, y_{i2}) & \dots & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & -(y_{i0}, y_{i1}, \dots, y_{iT-2}) \\ 0 & 0 & 0 & \dots & (y_{i0}, y_{i1}, \dots, y_{iT-2}) \end{bmatrix} \tag{1.46}$$

and the moment conditions in the first condition of (1.42). The moment condition in (1.43) leads to the moment condition being linear in ξ and can be written as $E(B_{1i}' u_i(\xi))$=0, where B_{1i} is the $T \times (T-2)$ matrix defined by

$$B_i = \begin{bmatrix} -y_{i1} & 0 & \dots & 0 \\ (y_{i1} + y_{i2}) & -y_{i2} & \dots & 0 \\ -y_{i2} & (y_{i2} + y_{i3}) & \dots & 0 \\ 0 & -y_{i3} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & -y_{i,T-2} \\ 0 & 0 & \dots & (y_{iT-2} + y_{i,T-1}) \\ 0 & 0 & \dots & -y_{i,T-1} \end{bmatrix} \tag{1.47}$$

The third of moment condition (1.42) will lead to moment condition being quadratic in ξ. Let S_i ($T \times l_1$) be made up of columns of R_i, A_i, and B_i, so that it represents some or all of the available linear instruments. The corresponding linear moment conditions are $E(f_i(\xi))$=0, with

$$f_i(\xi) = S_i'u_i(\xi) = f_{1i} + f_{2i}\xi, \quad f_{1i} = S_i'y_i, \; f_{2i} = -S_i'W_i. \tag{1.48}$$

Let the dimension of ξ be k, $l_1 \geq\ k$ is assumed such that f_i can identifyξ. The rest of $l_2 = l - l_1$ moment conditions will be written as $E(g_i(\xi))$=0. $g_i(\xi) = g_{1i} + g_{2i}\xi + (I \bigotimes \xi')g_{3i}\xi$, where g_{1i}, g_{2i}, and g_{3i} are $l_2 \times l$, $l_2 \times k$ and $l_2k \times k$ matrices in respective and the dimension of the identity matrix is l_2. $\hat{\xi}_{GMM}$ can be obtained by GMM based on all of the moment conditions:

$$E(m_i(\xi)) = E\begin{pmatrix} f_i(\xi) \\ g_i(\xi) \end{pmatrix} = 0$$

$m_N \equiv N^{-1}\sum_i m_i(\xi)$, with f_N, f_{1N}, f_{2N}, g_N, g_{1N}, g_{2N} and g_{3N} are defined similarly; $M_N \equiv \partial m_N/\partial\xi' \equiv\ (F_N', G_N')$ Let $M \equiv\ \operatorname*{plim}_{N\to\infty} M_N$, $F \equiv\ \operatorname*{plim}_{N\to\infty} M_N$, $F \equiv\ \operatorname*{plim}_{N\to\infty}\ F_N$ and $G \equiv\ \operatorname*{plim}_{N\to\infty}\ G_N$. Define the optimal weighting matrix:

$$\Omega \equiv \begin{bmatrix} \Omega_{ff} & \Omega_{fg} \\ \Omega_{gf} & \Omega_{gg} \end{bmatrix} \equiv \operatorname*{plim}_{N\to\infty} cov\left(\sqrt{N}\sum_i m_i\right).$$

Let $\hat{\Omega}$ be a consistent estimate of Ω of the form

$$\hat{\Omega} = \frac{1}{N}\sum_i m_i(\hat{\xi})m_i(\hat{\xi})',$$

where $\hat{\xi}$ is an initial consistent estimate of ξ. The efficient GMM estimator $\tilde{\xi}_{GMM}$ minimizes $N\ m_n(\xi)'\hat{\Omega}^{-1}m_N(\xi)$. The asymptotic covariance matrix of $N^{1/2}(\tilde{\xi}_{GMM} - \xi)$ is $[M'\Omega^{-1}M]^{-1}$. Ahn and Schmidt showed that the statistic $J_N = Nm_N(\tilde{\xi}_{GMM}'\hat{\Omega}^{-1}m_N\tilde{\xi}_{GMM})$ can be used to test the validity of the moment condition $E(m_i(\xi))$=0. The statistic is asymptotically chi-squared with (l-k) degrees of freedom under the joint hypothesis that all the moment conditions are valid.

It is worth noticing that the assumptions of uncorrelatedness (i) - (iv) is not sufficient to imply the asymptotic equivalence of IV and GMM estimates when the moment conditions of the first condition on (1.42) because the asymptotic equivalence of IV and GMM includes the restrictions on the fourth moment condition (such as $cov(y_{i0}^2, \varepsilon_{it}^2)$). Ahn (1990) has shown that the asymptotic equivalence of IV and GMM based on the first condition on (1.42) is ensured when (i) - (iii) are strengthened by independence assumption while (iv) is maintained. Wooldridge (1996) proposed more general treatment of case in which the asymptotic equivalence of IV and GMM holds if we replace the assumptions (i) - (iv) with conditional expectations as follows:

$$\begin{aligned} E(\varepsilon_{it}|y_{i0}, a_i, \varepsilon_{i1}, \ldots, \varepsilon_{i,t-1}) &= 0 \\ E(\varepsilon_{it}^2|y_{i0}, a_i, \varepsilon_{i1}, \ldots, \varepsilon_{i,t-1}) &= \sigma_{\varepsilon\varepsilon}. \end{aligned} \tag{1.49}$$

The conditions of (1.49) will be employed in the study of the conditional maximum likelihood estimator in later chapters.

Blundell and Bond (1998) consider two alternative estimators that impose further restrictions on the initial conditions process, designed to improve the properties of the standard-difference GMM estimator. The one is extended linear GMM estimator that uses lagged differences of y_{it} as instruments for equations in level, in addition to lagged levels of y_{it} as instruments for equations in first difference. The other is the use of the error components GLS estimator on an extended model that conditions on the observed initial values. Both estimators require restrictions on the initial conditions process. Asymptotic efficiency comparisons and Monte Carlo simulations for the simple AR(1) model demonstrate the dramatic improvement in performance of the proposed estimator compared to the usual first-difference GMM estimator, and compared to non-linear GMM. They use stationarity-like assumptions to show that system GMM estimators work better than the other estimator for ρ close to 1 by exploiting all moment conditions: $E(y_{i,t-s}\Delta\varepsilon_{it}) = 0$ for t=3,..., T and $s \geq 2$, $E(u_{i,t}\Delta y_{i,t-1})$=0 for t=4,5,..., T and $E(u_{i,3}\Delta y_{i2})$=0 as well as $E(y_{it}u_{it} - y_{i,t-1}u_{i,t-1})$ coming from homoscedasticity restrictions on ε_{it} for t=3 ,..., T.

Hahn (1999) consider the AR(1) panel model with fixed-effects. He investigate the estimation method developed by Blundell and Bond (1998), which makes use of the stationarity of the initial levels. By semi-parametric methods, he investigate an alternative linear GMM estimator based on additional moment restrictions, which are valid if we have

$$y_0 = \frac{a}{1-\rho} + u = \frac{a}{1-\rho} + \sum_{t=0}^{\infty} \rho^t \beta \varepsilon_t,$$

where ε_t are i.i.d. mean-zero random variables. By numerically comparing the semi-parametric information bounds for the case that incorporates the stationarity of the initial condition and for the case which does not, it is found that the efficiency gain is potentially.

Im, Ahn, Schmidt and Wooldridge (1999) showed that with panel data, exogeneity assumptions imply many more moment conditions than standard estimators use. However, many of the moment conditions may be redundant can not increase efficiency. They propose to establish the standard estimators' efficiency. The redundancy of moment conditions in GMM depends on relationships between the matrix of expected derivatives of the moment conditions and the optimal weighting matrix. They established results under assumption of no conditional heteroscedasticity, which implies a simple and tractable from for the optimal weighting matrix. They prove efficiency results for GLS in a model whit unrestricted error covariance matrix, and for 3SLS in a models where regressors and errors are correlated, for example the Hausman-Taylor model. For models with correlation between regressors and errors, and with unrestricted error covariance structure, they provide a simple estimator based on a GLS generalization of deviations from means (see Im, Ahn, Schmidt and

Wooldridge 1999).

1.6 Conclusion

The presence of correlation between the initial observations and unobserved effects contaminates the consistency of within estimator and related methods. Many approaches suggest a transformation to remove the unobserved effects, and then choose instruments based on sequential conditional moment assumptions.(for example, Anderson and Hsiao [1981]; Hsiao [1986]; Holtz and Eakin [1988]; Holtz, Eakin, Newey, and Rosen [1988]; Arellano and Bover [1990]; Arellano and Bond [1991].) While this treatment leads to the consistent estimators, the estimators are not efficient under standard assumptions because it does not make use of all of the available moment conditions. The random-effects formulation raises the interpretation of initial values. The traditional ML estimator (Balestra and Nerlove [1966]) is not generally consistent when y_{i0} is allowed to be random.

This question will not occur when y_{i0} is included in the joint density function as in unconditional ML estimation. When the other exogenous variables are introduced into the form of y_{i0}, the ML estimation of model (1.30) can be very complex. Sevestre and Trognon (1990) proposed another auxiliary autoregressive regression to tackle this problem. These treatments of initial conditions appears inflexible. Especially, when the non-linear model is necessary, it is difficult to find or approximate a proper distribution of y_{i0}. The different treatment of initial conditions will be introduced in subsequent chapters.

Although the λ-class estimation is often not practical, the asymptotic bias of λ-class estimator sheds some light on the existence of more efficient estimators. By containing the additional moment conditions identified in the Ahn and Schmidt's paper, the extended GMM leads to nontrivial gains in asymptotic efficiency. Several literatures suggest that imposing moment restrictions in GMM estimation to obtain more efficient estimators over a certain range of parameter space. Blundell-Bond (1998) and Hahn (1999) have shown the improvement on the efficiency is significant by the inclusion of the initial conditions.

Wooldridge (2000b) proposes a different treatment of initial conditions. His suggestion is to model $D(a_i|y_{i0}, X_{iT})$ and then construct the density of $(y_{iT}, \ldots, y_{i1})$ given (y_{i0}, X_{iT}). This allows y_{i0} to be random and does not require us to find, or even approximate $D(y_{i0}|a_i, X_{iT})$. Further, we need not specify an additional model for $D(a_i|X_{iT})$, or assume that a_i and X_{iT} are independent and then model $D(a)$. According to this framework, we can easily construct the conditional log likelihood function to obtain the consistent conditional maximum likelihood estimator. According to this framework, we can easily construct the conditional log likelihood function to obtain the consistent conditional maximum likelihood estimator. It pays to inefficiency that the method, conditional maximum likelihood estimator (CMLE), for handling the initial conditions problem appears to be novel, and offers a flexible, relatively simple alternative to the previous ones. The following chapters applies the CMLE to deal with some topics of dynamic panel data models: linear AR(1) dynamic panel data model with unobserved effect, the state dependence interacting with the unobserved effects, and the binary choice model.

Table 1.1: Consistency Properies of the MLEs for Dynamic Unobserved-effect Models

Case		(A)	(B)
y_{i0} fixed	ρ, β, σ_u^2	*Consistent*	*Consistent*
	γ, σ_α^2	*Inconsistent*	*Consistent*
y_{i0} random			
(a)y_{i0} independent of a_i	ρ, β, σ_u^2	*Consistent*	*Consistent*
	$\mu_{y0}, \gamma, \sigma_\alpha^2, \sigma_{y0}^2$	*Inconsistent*	*Consistent*
(b)y_{i0} independent of a_i	ρ, β, σ_u^2	*Consistent*	*Consistent*
	$\mu_{y0}, \gamma, \sigma_\alpha^2, \sigma_{y0}^2, \phi$	*Inconsistent*	*Consistent*

Case (A) is N fixed , T $\rightarrow \infty$
Case (B) is T fixed , N $\rightarrow \infty$

CHAPTER 2

Conditional Maximum Likelihood Estimator For The AR(1) Model

2.1 Introduction

A panel data model allows us to study the dynamics of economic behavior at an individual level in which the individual heterogeneity is taken into consideration. As discussed in Chapter 1, fixed effects approach does not lead to a consistent estimates for the parameters. The inconsistency mainly comes from the fact that the within transformation induces a correlation of order $\frac{1}{T}$ between variable and the error. The estimator of the conditional MLE explored by Balestra and Nerlove (1966) is not consistent as well because they treat the initial observations as nonrandom and such estimators, for a wide of combinations of the parameters, are equal to the within estimator and thus they are not consistent (see Trognon [1978]). Moreover, it is an untenable assumption to treat the first observation as nonrandom, since that implies it is independent of any other exogenous variables and any unobserved heterogeneity.

It is a natural solution to the estimation problem is to use maximum likelihood principle when the disturbances are assumed to be normal. The assumption of fixed initial observations can be relaxed when the likelihood function takes into consideration the density function of the initial observations, that is, the likelihood function is "unconditional" (Barghava and Sargan [1983]). In such an approach, we first describe the distribution of the dependent variables $(y_T, \ldots, y_0)$ conditional on $(x_T, \ldots, x_1, a)$, where x_t is strictly exogenous variables and a individual heterogeneity, t=1,...,T. We can specify the distribution of y_0 given (X_T, a) to obtain the distribution,$D(y_T, \ldots, y_0|X_T, a)$ and then integrate out a by specifying $D(a|X_T)$ or more typically just assuming that a is independent of X_T. This leads to a parametric density function $f(y_T, \ldots, y_0|X_T; \theta_0)$, which allows us to obtain the conditional maximum likelihood estimation with conditioning on X_T. Traditionally, this is viewed as "unconditional" MLE because the X_T are treated as nonrandom. Unfortunately, such an approach is made possible only provided that we have a steady state distribution for y_{it}. The inclusion of X_T makes matters more complicate. Barghava and Sargan (1983) treated the initial observations as random accounted for by time-constant variables and random errors in (1.31) as follows:

$$y_{i0} = \phi z_i + \nu_{i0},$$

where z_i is time-constant variable. The unobserved effects is assumed to be (1.32):

$$a_i = \psi \nu_{i0} + c_i,$$

where c_i is independent of ν_{i0}. Such a framework has been used, for example by Chamberlain (1984) and Blundell and Smith (1991). This can be solved by two-step way we have discussed in Chapter 1. However, this setup for initial observation do not include the exogenous variables. Sevestre and Trognon (1990) added another term

$x_{i0}\alpha$ into (1.31) meanwhile he need do estimate the autoregressive auxiliary model (1.36) beforehand. This method leads to a two-step estimation.

In the non-linear case (e.g. dynamic probit model), Heckman (1981) first make approximation to $D(y_0|X_T, a)$ and specify a $D(a)$ with assuming that a and X_T are independent. This method is flexible but it is more complicated and more restrictive than necessary. The misspecification of the distribution of y_{i0} would result in the inconsistency of the resultant estimator. It is obvious that the consistency properties of various error component estimators for the dynamic models with unobserved effects depends on the treatment with the initial value. Different assumption on the initial value induce more moment conditions needed to be exploited to gain more efficiency (e.g. Ahn and Schmidt [1995, 1997], Blundell and Bond [1998]).

The important drawbacks of unconditional MLE do not occur when we consider the distribution of $(y_T, \ldots, y_1)$ given (y_0, X_T, a) and then specify D$(a|y_0, X_T)$. This leads directly to a density for $(y_T, \ldots, y_1)$ given (y_0, X_T). Moreover, we do not treat y_0 as nonrandom variable and it in not necessary to assume the independence between X_T and a. Our suggestion is to model $D(a|y_0, X_T)$ and then construct the density of $(y_T, \ldots, y_1)$ given (y_0, X_T, a). This allows us to avoid the problem of having to find or even approximate, $D(y_0|X_T, a)$ and specify an auxiliary model for $D(c|X_T)$ or assume that a and X_T are independent and then model a marginal distribution of a (See Wooldridge [2000b]).

In this chapter, I first show how to construct the conditional MLE for

$$y_{it} = \rho y_{i,t-1} + a_i + \varepsilon_{it}, \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, T, \tag{2.1}$$

where the a_i is the individual effect and is assumed that $a_i = \alpha_0 + \alpha_1 y_{i0} + c_i$. ε_{it} and c_i are assumed to be normally distributed. Later on I consider the case with exogenous variables in which equation (2.1) will be added by the term $x_{it}\beta$ and that of a_i will be altered by adding one more term, $\bar{x}_i\alpha_2$.

The approach of CMLE keeps us away from understanding the exact form of the distribution of the first observation because it is conditioned on the initial observation. To specify an auxiliary conditional distribution for the unobserved heterogeneity has inherent drawback of all parametric methods: misspecification of this distribution generally results in inconsistent parameter estimates. Nevertheless, Wooldridge (2000b) has shown that in some leading cases the method leads to some remarkably simple conditional maximum likelihood estimators (especially for the non-linear case: partial effects on the mean response, averaged across the population distribution of the unobserved heterogeneity). For example, it is easy to obtain estimated average probability response across the population distribution of the unobserved heterogeneity discussed in Chapter 4. The plan of this chapter is as follows. Section 2 considers the general conditional MLE for the dynamic model. In this section I construct the conditional likelihood function to obtain the conditional maximum likelihood estimators and discuss the consistency of CMLE. Section 3 applies the CMLE to basic AR(1) model with unobserved effects. I examine the asymptotic properties of the CMLE as $N \rightarrow \infty$ with fixed T. Beginning with normality assumption on the unobserved effects and the random noises, I examine the AR(1) regression of dependent variables without exogenous variables and conduct a Monte Carlo studies to investigate the performance of the conditional maximum likelihood estimator. Theoretically, non-normality is known not to cause inconsistent in Gaussian CMLE. I proceed with the same studies with the replacement of normality by non-normality assumption. Section 4 examines the same model except that we include the strictly exogenous variables and employs the same procedure as that of section 3 to build up a simulation for the CMLE with inclusion of exogenous variables. Section 5 studies some empirical example for the previous two case. Section 6 makes the comparison of CMLE with the estimators discussed by Blundell and Bond (1998). Section 7 contains some concluding remarks.

2.2 General CMLE

2.2.1 Conditional Likelihood Function

In this section I will construct a generic likelihood function for the conditional maximum likelihood estimator in dynamic, unobserved effects models where the lagged value of dependent variable is included in the list of explanatory variables. The AR(1) model is a good choice to describe such a dynamic process. The primary principle on which estimation will be based is maximum likelihood. Let θ denote the vector of population parameters. Suppose we have observed a sample of size T+1, $(y_0, y_1, \ldots, y_T)$. We need find a joint distribution of $D(y_T, \ldots, y_1|y_0, X_T, a)$ where a is unobserved heterogeneity and its relevant parameterizing joint density function conditional on (y_0, X_T, a) is $f(y_T, \ldots, y_1|y_0, X_T, a; \delta)$ and thus the MLE estimate of θ is the value for which this sample is most likely to have been observed. Because a is unobserved, we need try to remove it out of the function. Typically, a distribution $D(a|y_0, X_T)$ is required and hence we can integrate a out of the joint density function with conditioning on X_T and y_0 by the usual product law. We make some assumptions in the following.

$$D(y_t|x_t, Y_{t-1}, a) = D(y_t|X_T, Y_{t-1}, a), \tag{2.2}$$

The assumption of (2.2) can be thought of as a basis for a standard dynamic unobserved effects analysis with strictly exogenous variables that means that, once current

x_t, past y_t and a are controlled for,x_s , $s \neq t$, has no effect on the distribution of Y_t. Therefore, we can define a parameterizing density for the conditional distribution of (2.2) as follows:

$$f_t(y_t|Y_{t-1}, x_t, a; \delta_0), \quad t = 1, \dots, T. \tag{2.3}$$

According to (2.3). the joint density of first t observations can be described as the product of $f(y_s|Y_{s-1}, x_s, a; \delta_0)$ over 1 to t. It follows that the parametric density of $(y_T, \dots, y_1)$ given (y_0, X_T, a) is

$$f(y_T, \dots, y_1|y_0, X_T, a; \delta_0) = \prod_{t=1}^{T} f_t(y_t|Y_{t-1}, x_t, a; \delta_0). \tag{2.4}$$

To integrate a out of the density function, Wooldridge (2000b) suggests modeling $D(a|y_0, X_T)$. We define the parametric density function as follows:

$$h(a|X_T, y_0; \lambda_0) \tag{2.5}$$

corresponding to $D(a|X_T, y_0)$, where λ is a vector of parameters. For example, we can assume that $E(a|y_0, X_T) = \mu(y_0, W_i)$, where W_i might be some linear combination of X_T. The simple case is that $W_i = \bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$. By the usual product law for conditional densities, the joint parametric density of $y_T, \dots, y_1$ given (y_0, X_T, a) is

$$p(y_T, \dots, y_1|y_0, X_T, a; \theta_0) = \prod_{t=1}^{T} f_t(y_t|Y_{t-1}, x_t, a; \delta_0) h(a|y_0, X_T; \lambda_0). \tag{2.6}$$

where $\theta_0 = (\delta_0, \lambda_0)$. Once we have specified $h(a|y_0, X_T; \lambda_0)$, we obtain the log density of $(y_T, \dots, y_1)$ given (y_0, X_T) by integrating out a.

$$\log \int_{\mathbb{R}^m} f(y_T, \dots, y_1|y_0, X_T; \delta_0) h(a|X_T, y_0; \lambda_0) v(\mathrm{d}\, a), \tag{2.7}$$

where m is the dimension of a and $v(\cdot)$ is a suitable measure. We let m = 1 and begin with the normality where the conditional mean, and possibly the conditional variance, are flexible functions of (y_0, X_T). For example, we assume that

$$a_i = \alpha_0 + \alpha_1 y_{i0} + c_i \tag{2.8}$$

when we consider the standard dynamic panel data without exogenous variables. It is usually assumed that $c_i|y_{i0} \sim \mathrm{N}(0, \sigma_a^2)$. It means that $h(c_i|y_{i0}, \lambda) = \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp(-\frac{1}{2}(\frac{c_i}{\sigma_a})^2)$, where $c_i = a_i - \alpha_0 - \alpha_1 y_{i0}$ and $\lambda = (\alpha_0, \alpha_1, \sigma_a^2)$. Given (2.6) and (2.7) without X_T, we can build up the log-likelihood function for the model without exogenous variables for cross section i is

$$\begin{aligned} & l(y_{iT}, \ldots, y_{i0}; \theta) \\ & = \log \int_{\mathbb{R}} f(y_{iT}, \ldots, y_{i1}|y_{i0}, a; \delta) h(c|y_{i0}; \lambda) \, \mathrm{d}\, c \\ & = \log \int_{\mathbb{R}} \textstyle\prod_{t=1}^{T} f_t(y_{it}|Y_{i,t-1}, a_i; \delta) h(c|y_{i0}; \lambda) \, \mathrm{d}\, c \end{aligned} \tag{2.9}$$

According to (2.9), we maximize the sum of $l\ (y_{iT}, \ldots, y_{i0}; \theta)$ across i from 1 to N. The log-likelihood is as follows:

$$\max_{\theta} \sum_{i=1}^{N} \log \int_{\mathbb{R}} \prod_{t=1}^{T} f_t(y_{it}|Y_{i,t-1}, a_i; \delta) h(c|y_{i0}; \lambda) \, \mathrm{d}\, c. \tag{2.10}$$

To extend the model to include strictly exogenous variables, the simple case is to specify an equation as follows:

$$a_i = \alpha_0 + \alpha_1 y_{i0} + \overline{x}_i \alpha_2 + c_i, \tag{2.11}$$

where $c_i|y_{i0}, x_i$ is Normal(0, σ_a^2). The setup of the model with strictly exogenous variables is as follows:

$$\max_{<\theta>} \sum_{i=1}^{N} \log \int_{\mathbb{R}} \prod_{t=1}^{T} f_t(y_{it}|Y_{i,t-1}, x_{it}, a_i; \delta) h(c|y_{i0}, x_i; \lambda) \, \mathrm{d}\, c, \tag{2.12}$$

where $\lambda = (\alpha_0, \alpha_1, \alpha_2, \sigma_a^2)$ and $h(c_i|y_{i0}, x_i; \lambda) = \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp(-\frac{1}{2}(\frac{c_i}{\sigma_a})^2)$, where $c_i = a_i - \alpha_0 - \alpha_1 y_{i0} - \overline{x}_i \alpha_2$.

A different description of the likelihood function for a sample of size T from a Gaussian AR(1) with unobserved effects is sometimes useful. Let $(y_i|y_{i0}, x_i, a) = (y_{iT}, \ldots, y_{i1}|y_{i0}, x_i, a_i)$ could be viewed as a single realization given (y_{i0}, x_i, a_i) from a

T-dimensional Gaussian distribution. Viewing the observed sample y_i as a single draw from a Normal($\mu(y_{i0}, x_i)$, $\Omega(y_{i0}, x_i)$) where $\mu(y_{i0}, x_i) = E(y_i|y_{i0}, x_i, a_i)$ and $\Omega(y_{i0}, x_i) = Var(y_i|y_{i0}, x_i, a_i)$, the sample likelihood function could be written down from the formula for the multivariate Gaussian density:

$$f(y_i; \delta) = (2\pi)^{\frac{-T}{2}} (|\Omega(y_{i0}, x_i)|^{\frac{-1}{2}} \exp[-\frac{1}{2}(y_i - \mu(y_{i0}, x_i))'\Omega(y_{i0}, x_i)^{-1}(y_i - \mu(y_{i0}, x_i))]. \tag{2.13}$$

By specifying a Gaussian distribution $h(a_i|y_{i0}, x_i)$, the individual log likelihood function can be written as follows:

$$\log \int_{\mathbb{R}^m} (2\pi)^{\frac{-T}{2}} (|\Omega(y_{i0}, x_i)|^{\frac{-1}{2}} \exp[-\tfrac{1}{2}(y_i - \mu(y_{i0}, x_i))'\Omega(y_{i0}, x_i)^{-1}(y_i - \mu(y_{i0}, x_i))] \cdot h(a|y_{i0}, x_i) \mathrm{d}\, a. \tag{2.14}$$

Expression (2.13) is algebraically equal to (2.7). We can maximize the sum of (2.14) with respective to θ across i from 1 to N to obtain the CMLE estimators.

2.2.2 Asymptotic Properties of the CMLE

In the current setting, the conditional maximum likelihood estimator is generally consistent – with fixed T and N goes to infinity – if the conditional density of $(y_{i1}, \ldots, y_{iT})$ given (y_{i0}, X_{iT}) is correctly specified. This follows from standard results on maximum likelihood estimation with conditioning variables because we are assuming random sampling in the cross section.(See, for example, Manski(1988, Chapter 5)), Wooldridge (2001, Chapter 13).) In the present application to linear, dynamic unobserved effects models, the log-likelihood function satisfies all smoothness requirements, and the sufficient moment conditions are likely to be met. Practically, the key issue is parameters are identified under weaker assumptions based only on certain moment conditions, so identification holds when we specify a full conditional distribution.)

It is useful to sketch the consistency of the CMLE for general dynamic models where the likelihood function is conditional on the initial value. The density in 2.7)is correctly specified if there are values δ_0 and λ_0 such that the density of $(y_{i1}, \ldots, y_{iT})$given (y_{i0}, X_{iT}) is given by the integral in (2.7). Under this assumption, the conditional Kullback-Leibler information inequality holds (see, for example, Manksi (1988, Section 5.1)):

$$E(l(y_i;\theta_0)\,|x_i, y_{i0})) \geq E(l(y_i;\theta)\,|x_i, y_{i0})), \tag{2.15}$$

for all θ in the parameter space. By the law of iterated expectations and (refeq2-10) we have

$$E\left[l(y_i, x_i;\theta_0)\right] \geq E\left[l(y_i, x_i;\theta)\right]. \tag{2.16}$$

Therefore, θ_0 is a solution to the population maximization problem:

$$\max_{\theta\in\Theta} E\left[l(y_i, x_i;\theta)\right]. \tag{2.17}$$

This shows that the CMLE is Fisher consistent for θ_0. Under identification, θ_0 is the unique solution to (2.15). Then, we can use the usual analogy principle and the uniform weak law of large numbers to conclude that the CMLE is generally consistent for θ_0 as $N \to \infty$.

In rare situations, the log-likelihood function can be shown to be globally concave. Unfortunately, this does not appear to be the case for dynamic panel data models. As a practical matter, this means we may locate local extrema. In practice, several different starting values should be used in estimation to try to uncover a global maximum.

Under sufficient differentiability assumptions - which, as mentioned earlier, are satisfied by the models of this and the remaining chapters - the CMLE is $\sqrt{N}$ asymptotically normal. Newey and McFadden(1994) and Wooldridge(2001) show that a consistent root to the maximization problem is also asymptotically normal:

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \rightarrow \mathrm{N}(0,\ \mathrm{A}(\theta_0)^{-1}\mathrm{B}(\theta_0)\mathrm{A}(\theta_0)^{-1}),$$

where

$$A(\theta_0) = E[(\partial^2 l(y_i;\theta)/\partial\theta\partial\theta')_{\theta_0}],$$

and

$$B(\theta_0) = E[(\partial l(y_i;\theta)/\partial\theta)_{\theta_0} \times (\partial l(y_i;\theta)/\partial\theta')_{\theta_0}].$$

Under correct specification of the conditional density, $A(\theta_0) = -B(\theta_0)$, that is, the information matrix equality holds. This simplifies estimation of the asymptotic variance and computation of test statistics.

2.3 Linear AR(1) Model With Unobserved Effects

2.3.1 Linear AR(1) Model

The conditional MLE approach is one method for making the initial condition problem tractable. We begin with the linear case without additional explanatory variables. The model is

$$y_{it} = \rho\, y_{i,t-1} + a_i + \varepsilon_{it}, \tag{2.18}$$

and we make the following assumptions.

Assumption 2.1 $\varepsilon_{it}|y_{i,t-1},\ldots,y_{i0},a_i \sim \mathrm{Normal}(0,\sigma_\varepsilon^2)$.

Assumption 2.2 $a_i|y_{i0} \sim \mathrm{Normal}(\alpha_0 + \alpha_1 y_{i0}, \sigma_a^2)$.

According to Assumption 2.1 and Assumption 2.2, the distribution for

$(y_{iT}, y_{i,T-1}, \dots, y_{i1})$ conditioning on y_{i0} is as follows:

$$(y_{iT}, y_{i,T-1}, \dots, y_{i1} |\, y_{i0}) \sim \mathrm{N}(\mu(\mathrm{y_{i0}}), \Omega(\mathrm{y_{i0}})), \quad \mathrm{i} = 1, \dots, \mathrm{N}, \tag{2.19}$$

where

$$\begin{aligned} \mu(y_{i0}) &= E(y_i|y_{i0}), \\ \Omega(y_{i0}) &= V(y_{i0}|y_{i0}) \\ &= E\left((y_i - E(y_i|y_{i0}))(y_i - E(y_i|y_{i0}))'|y_{i0}\right) \\ & i = 1, \dots, N,\ y_i = (y_{iT}, \dots, y_{i1}). \end{aligned}$$

Assumption 2.2 implies that

$$a_i = \alpha_0 + \alpha_1 y_{i0} + c_i,\ where\ c_i|y_{i0} \sim \mathrm{N}(0, \sigma_{\mathrm{a}}^2) \tag{2.20}$$

Equation (2.1) can be re-written as $y_{it} = \rho^t y_{i0} + \sum_{j=1}^t \rho^{j-1}(\alpha_0 + \alpha_1 y_{i0} + c_i) + \sum_{j=1}^t \rho^{j-1}\varepsilon_{i,t-j+1}$. The $E(y_i|y_{i0})$ can be obtained by replaceing $E(a_i|y_{i0})$ with $\alpha_0 + \alpha_1 y_{i0}$ in (2.19). The conditional mean of y_i is as follows

$$\mu(y_{i0}) = \left(\alpha_0 + (\alpha_1 + \rho)y_{i0}, \ \dots \ , \frac{1-\rho^t}{1-\rho}\alpha_0 + (\frac{1-\rho^t}{1-\rho}\alpha_1 + \rho^t)y_{i0}, \ \dots \right)' \tag{2.21}$$

The conditional variance of y_{it} can be obtained by calculating the form as $E(\varepsilon(y_{i0})\varepsilon(y_{i0})'|\, y_{i0})$, where $\varepsilon(y_{i0}) = y_i - E(y_i|y_{i0})$. In the same manipulation as that of conditional mean, the conditional variance can be written as

$$\Omega(y_{i0}) = \begin{pmatrix} \omega_{11} & \dots & \omega_{1T} \\ \vdots & \ddots & \vdots \\ \omega_{T1} & \dots & \omega_{TT} \end{pmatrix},$$

where

$$\begin{aligned} \omega_{tt} &= (\frac{1-\rho^t}{1-\rho})^2\sigma_a^2 + (\frac{1-\rho^{2t}}{1-\rho^2})\sigma_\varepsilon^2\ ; \qquad t = 1,\dots,T. \\ \omega_{st} &= (\frac{1-\rho^s}{1-\rho}\frac{1-\rho^t}{1-\rho})\sigma_a^2 + \rho^{|t-s|}(\frac{1-\rho^{2s}}{1-\rho^2})\sigma_\varepsilon^2\ ; \quad s \neq t\ ,\ s,t = 1,\dots,T. \end{aligned} \tag{2.22}$$

The jointly parametric density function of $y_i|y_{i0}$:

$$f(y_i|y_{i0};\theta) = (\frac{1}{\sqrt{2\pi}})^{-T/2}(|\Omega(y_{i0})|)^{-1/2}\exp(\frac{-1}{2}\left(\varepsilon(y_{i0})'\Omega(y_{i0})^{-1}\varepsilon(y_{i0})\right)) \tag{2.23}$$

where $\varepsilon(y_{i0}) = y_i - E(y_i|y_{i0})$. We can directly construct the log-likelihood function across i from 1 to N as follows:

$$\mathcal{L}(Y;\theta) = \sum_{i=1}^{N}\left(\frac{-T}{2}\log\sqrt{2\pi} + \frac{1}{2}\log|\Omega(y_{i0})|^{-1} - \frac{1}{2}\{\varepsilon(y_{i0})'\Omega(y_{i0})^{-1}\varepsilon(y_{i0})\}\right), \tag{2.24}$$

where $\theta = (\rho, \alpha_0, \alpha_1, \sigma_a^2, \sigma_\varepsilon^2)$. The CMLE estimators can be obtained by maximizing the likelihood function (2.24).

Another approach to calculate the CMLE estimators, according to equation (2.10) and Assumption 2.2, is in the following. We specify the distribution of a_i conditioning on y_{i0} as follows:

$$h(a_i|y_{i0};\lambda_0) = \frac{1}{\sqrt{2\pi\sigma_a^2}}\exp(\frac{-1}{2}(\frac{a_i - \alpha_0 - \alpha_1 y_{i0}}{\sigma_a})^2). \tag{2.25}$$

By employing (2.13), the joint density of y_i given (y_{i0}, a_i) is the product of $f(y_{iT},\ldots,y_{i1}|a_i,y_{i0})$ and $h(a_i|y_{i0})$, where the f(·) and h(·) are the relevant conditional normal density functions. It follows that the density of $(y_{iT},\ldots,y_{i1})$ given $(y_{i0};\theta)$ is

$l_i(y_i;\theta) =$

$$\log\int_{-\infty}^{\infty}(\frac{1}{\sqrt{2\pi}})^{-T/2}(|\Omega(y_{i0})|^{-1/2}\exp[-\frac{1}{2}(y_i-\mu(y_{i0}))'\Omega(y_{i0})^{-1}(y_i-\mu(y_{i0}))]\cdot \frac{1}{\sqrt{2\pi\sigma_a^2}}e^{\frac{-1}{2}(\frac{c}{\sigma_a})^2}\,dc, \tag{2.26}$$

where $\mu(y_{i0}) = \rho y_{i,-1} + a_i l_T$. Therefore, the CMLE estimators are to solve out the problem of maximization as follows:

$$\max_\theta \mathcal{L}(Y;\theta) = \max\sum_{i=1}^{N} l_i(y_i,\theta) \tag{2.27}$$

where θ is the vector of parameters. Because we place no restrictions on $\text{h}(a|y_{i0}, \lambda_0)$, once we have specified $h(a_i|y_{i0}; \lambda_0)$, we generally obtain the density of y_i given y_{i0} by integrating out a_i.

While the log-likelihood function is "consistent" under the normality for ε_{it} and a_i - and therefore, y_i given y_{i0} is multivariate normal - the conditional MLE is robust to ce'teris pa'ripus from the assumptions. In particular, the normal quasi-MLE is consistently and asymptotically normal provided the first two conditional moments, $E(y_i|y_{i0})$ and $Var(y_i|y_{i0})$ are correctly specified. this follows from the work of Gourieroux, Monfort and Trognon (1984) and Bollerslev and Wooldridge (1992). Without normality, the information matrix equally does not hold and so the variance matrix needs to be estimated in a robust way.

2.3.2 Simulation Evidence

In order to investigate the performance of maximum-likelihood estimators given the initial value, we conducted Monte Carlo studies. We use the MLE software of Gauss to do our simulation for the conditional maximum likelihood function. The notations for the simulation are as follows:

1. θ^* means the conditional maximum likelihood estimators in each iteration.

2. $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$.

3. θ means true value of parameter, where $\theta = (\rho, \alpha_0, \alpha_1, \sigma_a, \sigma_\varepsilon) =$

   $(\rho, 0.2, 0.4, \sqrt{1.2}, \sqrt{2.4})$.

Our true models were generated by

$$y_{it} = \rho \; y_{i,t-1} + a_i + \varepsilon_{it}$$
$$i = 1, \ldots, 250, t = 1, \ldots, 5, \qquad (2.28)$$
$$\rho = 0, 0.05, \ldots, 0.95,$$

where

$$a_i = 0.2 + 0.4 \; y_{i0} + c_i, \quad i = 1, \ldots, 250. \qquad (2.29)$$

We generated the c_i and ε_{it} by two cases, one from in dependently normal distribution, $\varepsilon_{it} \sim$ N(0,2.4) and $c_i \sim$ N(0, 1.2) and the other from a t-distribution with the freedom 6 and 10, in respective. In case where y_{i0} are treated with being given, we do not need pay attention to its distribution in our approach. For the simplicity, we generate y_{it} from a N(0, 1) or uniform distribution for convenience. The value of ρ goes from 0 to 0.95 in an increment of 0.05. We use the individual likelihood function (2.26) and (2.27) and then construct the framework of maximization to solve out estimators.

The specification for the distribution of $(a_i|y_{i0})$ in the use of the framework (2.25) is flexible. We can see the advantage of framework (2.25) in non-linear model, for example logit with unobserved effects model will be discussed in chapter 4; it, nevertheless, is heavy time-consuming in the maximization of the likelihood function (2.26) across i to N. We employ the Hermite integral formula as the approximation of the integral (see Butler and Moffitt [1982]). It is a good idea in the use of framework (2.26) to specify a more flexible distribution of the unobserved heterogeneity given the growing speed of CPU.

Table 2.1 reports the simulation result for the power test of the conditional maximum likelihood estimators, $H_0 : \theta = (\rho, 0.2, 0.4, \sqrt{2.4}, \sqrt{1.2})$. The true values of ρ range from 0 to 0.95 with the increment of 0.05. We repeat the same procedure of the

CMLE for 1200 times and calculate the frequence of the p-value greater than a certain level, 0.01, 0.05 or 0.10. Table 2.2 shows the simulation for the test of $H_0 : \rho = \rho_0$. For example, under the p-value is 0.01 and the true value of ρ is 0, the second row of Table 2.2 shows that the frequence of rejecting $H_0 : \rho = \rho_0$ in creases with ρ_0, namely, we can reject most of ρ_0, away from the true value, ρ. There are same results for the p-value, 0.05 or 0.10.

It is crucial to see that it is more powerful to do the hypothesis $H_0 : \rho = \rho_0$ when the true value of ρ closer to 0. For example, the true value of ρ is 0.75 and the power 0.01 in the second row of Table 2.2, the p-value of $H_0 : \rho = 0.90$ is 0.6183. Comparing with $\rho = 0.75$, the p-value of $H_0 : \rho = 0.15$ is 0.9050 when the true value of ρ is 0; our approach, obviously, for the hypothesis test of ρ performs well when the true value of ρ is getting closer to 0.

To examine the simulation for the model under non-normality, we generate the ε_{it} and c_i from the t-distribution with freedom 6 and 10, i.e. the parameters, $\sigma_\varepsilon = \sqrt{\frac{6}{6-2}}$ and $\sigma_a = \sqrt{\frac{10}{10-8}}$, respectively. We report the simulation results for the conditional maximum likelihood estimators in Tables 2.3 - 2.4. Table 2.4 shows the simulation for the test of $H_0 : \rho = \rho_0$, with ρ_0 ranging from 0 to 0.95 for $\rho = 0, 0.1, 0.05, \ldots, 0.95$. We obtain similar results of the model with normality assumption. The simulation support that the conditional maximum likelihood estimator perform very well. The model of interest is a regression model in which the lagged value of the dependent variable appears in the list of explanatory variables, it is crucial for the test of coefficient of the lagged dependent variable, $H_0 : \rho = 0$. Our approach supports that the CMLE is a good estimator. When the true value is closer to zero, the test is more significant. The sixth column of Table 2.3, the frequence of rejecting the $H_0 : \sigma_a = \sqrt{1.2}$ is larger, namely it is likely to be rejected in comparison with the σ_ε. By increasing N, the power of testing σ_a will be increase. We have discussed the properties of CMLE for dynamic models with individual-specific effects. In the next section, we study the same linear AR(1) model with unobserved effects and strictly exogenous explanatory variables.

2.4 Linear AR(1) Model With Unobserved Effects And Exogenous Regressors

2.4.1 Linear AR(1) Model With Exogenous Variables

In this section, we add exogenous variable, x_{it}, to model (2.1). The new model is

$$y_{it} = \rho y_{i,t-1} + x_{it}\beta + a_i + \varepsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{2.30}$$

where x_{it} is assumed to be strictly exogenous variable. The exogenous variables might be the variables of discrete value, e.g. some policy variables, status variable and the like, or variables of continuous value, e.g. years of education. We make some assumption as follows in this case:

Assumption 2.3 $\varepsilon_{it}|y_{i,t-1}, \ldots, y_{i0}, x_{i,T}, \ldots, x_{i1}, a_i \sim \text{Normal}(0, \sigma_\varepsilon^2)$.

Assumption 2.4 $a_i|x_{it}, \ldots, x_{i1}, y_{i0} \sim \text{Normal}(\alpha_0 + \alpha_1 y_{i0} + \overline{x}_i \alpha_2, \sigma_a^2)$.

and thus we have the equation

$$E(a_i|y_{i0}, x_i) = \alpha_0 + \alpha_1 y_{i0} + \overline{x}_i \alpha_2, \tag{2.31}$$

where $\overline{x}_i = \frac{1}{N}\sum_{t=1}^{T} x_{it}$ and $x_i = (x_{iT}, \ldots, x_{i1})$. In the empirical study in section 2.5, we let x_{it} be a union status variable; then $\overline{x}_i$ is the fraction of time in a labor union over the sample period. For example, if a worker had been in labor union for three years, e.g. 1981, 1983 and 1984 from 1981 to 1987, then the ratio,$\overline{x}_i$ is $\frac{3}{7}$. We can construct the conditional multivariate normal distribution for the model by adding the exogenous variable into (2.1) and (2.19) as follows:

$$y_{iT}, y_{i,T-1}, \ldots, y_{i1} |\; y_{i0}, x_i \;\sim\; \mathrm{N}(\mu(\mathrm{y_{i0}, x_i}), \Omega(\mathrm{y_{i0}, x_i})), \quad \mathrm{i} = 1, \ldots, \mathrm{N}, \tag{2.32}$$

where

$$\begin{aligned} \mu(y_{i0}, x_i) &= E(y_i | y_{i0}, x_i); \\ \Omega(y_{i0}, x_i) &= V(y_i | y_{i0}, x_i) \\ &= E\left((y_i - E(y_i|y_{i0}, x_i))(y_i - E(y_i|y_{i0}, x_i))' | y_{i0}, x_i\right) \\ i = 1, \ldots, N,\ y_i &= (y_{iT}, \ldots, y_{i1}),\ x_i = (x_{iT}, \ldots, x_{i1}). \end{aligned} \tag{2.33}$$

With Assumption 2.4, we rewrite (2.20) as follows

$$a_i = \alpha_0 + \alpha_1\, y_{i0} + \overline{x}_i\, \alpha_2 + c_i, \tag{2.34}$$

where $c_i | y_{i0}, x_{iT}, \ldots, x_{i1} \sim \mathrm{N}(0, \sigma_{\mathrm{a}}^2)$.

2.4.2 Conditional Mean and Variance

By iteration, equation (2.31) can be expressed as

$$y_{it} = \rho^t y_{i0} + \sum_{j=1}^{t} \rho^{j-1}(\alpha_0 + \alpha_1 y_{i0} + \overline{x}_i\, \alpha_2 + c_i) + \sum_{j=1}^{t} \rho^{j-1} x_{i,t-j+1} + \sum_{j=1}^{t} \rho^{j-1} \varepsilon_{i,t-j+1}. \tag{2.35}$$

The mean $E(y_i|y_{i0}, x_i)$ can be obtained by substituting $E(a_i|y_{i0}, x_i)$ with $\alpha_0 + \alpha_1 y_{i0} + \overline{x}_i \alpha_2$. The conditional mean of y_i is as follows

$$\mu(y_{i0}, x_i) = \begin{pmatrix} \alpha_0 + (\alpha_1 + \rho) y_{i0} + x_{i1} \\ \vdots \\ \frac{1-\rho^t}{1-\rho}(\alpha_0 + \overline{x}_i) + (\frac{1-\rho^t}{1-\rho}\alpha_1 + \rho^t) y_{i0} + \sum_{j=1}^{t} \rho^{t-j+1} x_{i,t-j+1} \\ \vdots \end{pmatrix} \tag{2.36}$$

The conditional variance of y_{it} can be obtained by calculating the form as equation (2.33) do. In the same manipulation as that of conditional mean, the conditional variance can be described as follows

$$\Omega(y_{i0}, x_i) = \begin{pmatrix} \omega_{11} & \dots & \omega_{1T} \\ \vdots & \ddots & \vdots \\ \omega_{T1} & \dots & \omega_{TT} \end{pmatrix}$$

Actually, provided that x_{it} is strictly exogenous, $\Omega(y_{i0}, x_i)$ is equal to $\Omega(y_{i0})$. Equations (2.23) and (2.24) can be applied here. We parameterize the conditional densities of $(y_i|y_{i0}, x_i)$:

$$f(y_i|y_{i0}, x_i; \theta) = (\frac{1}{\sqrt{2\pi}})^{-T/2}(|\Omega(y_{i0})|^{-1})^{1/2} \exp(\frac{-1}{2}(\varepsilon(y_{i0}, x_i))\Omega(y_{i0})^{-1}(\varepsilon(y_{i0}, x_i))'), \tag{2.37}$$

where $\varepsilon(y_{i0}, x_i) = y_i - E(y_i|y_{i0}, x_i)$. We can directly construct the log likelihood function across i from 1 to N as follows:

$$\mathcal{L}(Y, X; \theta) = \sum_{i=1}^{N} \left(\frac{-T}{2} \log \sqrt{2\pi} + \frac{1}{2} \log |\Omega(y_{i0})|^{-1} - \frac{1}{2} \left(\varepsilon(y_{i0}, x_i)'\Omega(y_{i0})^{-1}\varepsilon(y_{i0}, x_i)\right) \right), \tag{2.38}$$

where $\theta = (\rho, \beta, \alpha_0, \alpha_1, \sigma_a^2, \sigma_\varepsilon^2)$. The CMLE estimators can be obtained by maximizing the likelihood function (2.38).

Another approach to calculating the CMLE estimators, according to equation (2.10) and Assumption 2.4, is in the following. We specify the distribution of a_i conditioning on y_{i0}, x_i as follows:

$$h(a_i|y_i, x_i; \lambda_0) = \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp(\frac{-1}{2}(\frac{a_i - (\alpha_0 + \alpha_1\; y_{i0} + \overline{x}_i\; \alpha_2)}{\sigma_a})^2). \tag{2.39}$$

The likelihood function of this case is similar to that of the previous model without exogenous regressors except that the conditional mean, $\mu(y_{i0}) = E(y_i|y_{i0}, a_i)$, must be replaced with $\mu(y_{i0}, x_i) = E(y_i|y_{i0}, x_i, a_i)$, so equations (2.26) and (2.27) can be directly applied here. We write the likelihood function of interest as follows:

$l_i(y_i, x_i; \theta) =$

$$\log \int_{-\infty}^{\infty} (\frac{1}{\sqrt{2\pi}})^{-T/2}(|\Omega(y_{i0}, x_i)|)^{-1/2} \exp[-\tfrac{1}{2}(y_i - \mu(y_{i0}, x_i))'\Omega(y_{i0}, x_i)^{-1} \cdot (y_i - \mu(y_{i0}, x_i))] \frac{1}{\sqrt{2\pi\sigma_a^2}} e^{\frac{-1}{2}(\frac{c}{\sigma_a})^2} \, d\,c, \tag{2.40}$$

where $\mu(y_{i0}, x_i) = \rho y_{i,-1} + x_i\beta + a_i l_T$. Therefore, the CMLE estimators are to solve out the problem of maximization as follows:

$$\max_\theta \mathcal{L}(Y, X; \theta) = \max_\theta \sum_{i=1}^{N} l_i(y_i, x_i; \theta) \tag{2.41}$$

where θ is a vector of parameters. According to the framework discussed previously, I set up a simulation for it in section (2.4.3).

2.4.3 Simulation Evidence

I conducted Monte Carlo experiment to examine the performance of the CMLE model in which exogenous variables are included. I use the MLE software of Gauss to do the simulation for the conditional maximum likelihood estimator. The notations for the simulation are as follows:

1. $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200} \theta_j^*$, where θ_j^* is the estimates from the CMLE in each iteration.

2. θ means true value of parameter, where $\theta = (\rho, \beta, \alpha_0, \alpha_1, \alpha_2, \sigma_a, \sigma_\varepsilon) = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{1.2}, \sqrt{2.4})$. .

Our true model was generated by

$$y_{it} = \rho\, y_{i,t-1} + 0.15\, x_{it} + a_i + \varepsilon_{it}, \quad i = 1, \ldots, 250, \quad t = 1, \ldots, 5, \quad \rho = 0, 0.05, \ldots, 0.95. \tag{2.42}$$

where,

$$a_i = 0.2 + 0.4\, y_{i0} + 0.35\, \overline{x_i} + c_i, \quad i = 1, \ldots, 250. \tag{2.43}$$

I generated the c_i and ε_{it} by two ways, one from independently normal distribution, $\varepsilon_{it} \sim$ N(0,2.4) and $c_i \sim N(0, 1.2)$ and the other from t-distribution with the freedom 6 and 10, in respective. The value of ρ ranges from 0 to 0.95 with an increment of 0.05. I use the individual likelihood function (2.26) with replacement of $\mu(y_{i0}, x_i) = \rho y_{i,-1} + x_i\beta + a_i l_T$ and then construct

$$\max_{\theta} \sum_{i=1}^{N} l(y_i; \theta). \tag{2.44}$$

Table 2.6 reports the simulation results for the power of tests of $H_0 : \theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{2.4}, \sqrt{1.2})$. The true values of ρ range from 0 to 0.95 with the increment of 0.05. I repeat the same procedure of the CMLE for 1200 times and calculate the frequence of the p-value greater than a certain level, 0.01, 0.05 or 0.10. Table 2.7 shows the simulation for the test of $H_0 : \rho = \rho_0$. I calculate the frequency of p-value greater than a certain level, 0.01, 0.5 or 0.10. For example, under the p-value is 0.01 and the true value of ρ is 0, the second row of Table 2.7 shows that the frequence of rejecting $H_0 : \rho = \rho_0$ increases with ρ_0. It means that most of $\dot{\rho}_0$, away from the true value, ρ can be rejected in the CMLE. There are same results for the p-value, 0.05 or 0.10. Table 2.8 shows that the result of the simulation by replaceing the normality assumption with t distribution. Table 2.9 shows that the frequence of rejecting $H_0 : \rho = \rho_0$ increases with ρ_0 even without the normality assumption. It pays to notice the test of estimated standard deviation of unobserved effect when we drop the normality assumption. The 8th column of Table 2.8 $\sim$ Table 2.9, the frequence of rejecting the $H_0 : \sigma_a = \sqrt{1.25}$ is larger, namely it is likely to be rejected in comparison with the σ_ε. The reason might be that we generate the unobserved effects from the distribution from the t distribution, the variance will become larger and the number of a_i is much smaller than the number of ε_{it}. By increasing N value , the power of testing σ_a will be increased.

It is crucial to see that the true value of ρ getting closer to 0 or 1, it is more significant to reject $H_0 : \rho = \rho_0$ than the true value falling within the interval of 0 and 1 in which ρ_0 deviates from the true value. For example,let us set the deviation be three increment, 0.15, meaning the deviation is 0.05×3. The true value of ρ is 0.5 and the power 0.01 in the second row of Table 2.7, the p-value of $H_0 : \rho = 0.65$ is 0.7558; the p-value of $H_0 : \rho = 0.25$ is 0.8833 when the true value of ρ is 0.1 (see Table 2.7); the p-value of $H_0 : \rho = 0.8$ is 0.9908 when the true value of ρ is 0.95 (See Table 2.7).

The results of this simulation show that most of conditional maximum likelihood estimators deviating away from the true value of the associated parameter will be rejected, especially when the true value of parameter is getting closer to 0 or 1.

2.5 Empirical Example

I have discussed the properties of the conditional maximum likelihood estimators for dynamic model with individual heterogeneity in previous sections. In this section, I use the data from Vella and Verbeek (1998) to study the conditional maximum likelihood estimator in estimating dynamic model using observations draw from a time series of cross sections. These data are for young males taken from the National Longitudinal Survey (Youth Sample) for the period 1980 - 1987. The dependent variable is the log of hourly wage and the explanatory variable is labor union status. Each of the 545 men in the sample worked in every year from 1980 through 1987. We begin with the OLS for the empirical data, i.e. we run the OLS regression of $lnwage_t$ on 1, $lnwage_{i,t-1}$. The OLS estimates of autoregressive is 0.627. The OLS estimates cannot be identified with the effects of unobserved effects. It is necessary to incorporate the effect of individual heterogeneity to study both the state dependence in earnings as well as the effects of union status on wage. We assume that omitted ability and other productivity factors can be accounted for by initial wage rate. The model is set up as follows:

$$lnwage_{it} = \rho\ lnwage_{i,t-1} + \ a_i + \varepsilon_{it}, \quad i = 1, \ldots, 545, \quad t = 1, \ldots, 7, \tag{2.45}$$

where

$$a_i = \alpha_0 + \alpha_1\ lnwage_{i0} + \ c_i, \quad i = 1 \ldots, 545. \tag{2.46}$$

Table 2.5 shows the conditional maximum likelihood estimates, $(\hat{\rho},\ \hat{\alpha_0},\ \hat{\alpha_1})$= (0.3405, 0.8784, 0.1839) are all significantly different from zero. The estimated average effects of unobserved heterogeneity given initial log wage, $\hat{a}_i$ is measured by $(0.8784 + 0.1839\ lnwage_{i0})$. This verifies that the higher is the initial wage rate, the higher is the individual worker's ability. Replaceing a_i in (2.16) with the above equation and taking the mean of $lnwage_{it}$ given the $lnwage_{i0}$, equation (2.25) can be expressed as follows by iteration

$$E(lnwage_{it}|lnwage_{i0}) = \rho^t\ lnwage_{i0} + (1+\rho+\rho^2+\ldots+\rho^{t-1})\ (\alpha_0+\alpha_1\ lnwage_{i0}) \tag{2.47}$$

From equation (2.47), the estimated response of the current wage rate change into the initial wage, $\frac{\partial}{\partial\ lnwage_{it}} / \frac{\partial}{\partial\ lnwage_{i0}}$ is $(0.3405^t + 0.1839 \cdot \frac{1-0.3405^t}{1-0.3405})$ instead of 0.3405^t. Specifically, when t=1,2,...,7, the estimated responses are 0.4884, 0.5444,...,0.5688, respectively.

Vella and Verbeek (1998) study the effects of union membership on wages in a static model. Here I add union status to the AR(1) model with an unobserved effect. Specifically, the model is

$$lnwage_{it} = \rho\ lnwage_{i,t-1} + \beta\ union_{it} + \ a_i + \varepsilon_{it}, \quad i = 1, \ldots, 545, \quad t = 1, \ldots, 7, \tag{2.48}$$

where the unobserved effect is assumed to follow

$$a_i = \alpha_0 + \alpha_1\ lnwage_{i0} + \alpha_2\ \overline{union_i} + \ c_i, \quad i = 1 \ldots, 545. \tag{2.49}$$

Given past wage and controlling for unobserved heterogeneity, the return to union membership is about 4.7 percent and it is marginally statistically significant. The estimates suggest that, once the initial wage is controlled for, there is no partial correlation between individual heterogeneity and the propensity to belong to a union.

The analysis here assumes that union status is strictly exogenous. In the context of model (2.48), this means that innovations in lnwage today, as measured by ε_{it}, do not affect the decision to join a union in the future. This may not be true, although we are controlling already for the most recent wage and an unobserved effect. One way to test the strict exogeneity assumption is to put a lead of union, that is, $union_{i,t+1}$, in the equation and test its statistical significance.

The $\overline{union_i}$ is the ratio of periods staying in labor union to the periods outside of labor union for a given periods of time. For example, if a worker had been in labor union for three years, e.g. 1981, 1983 and 1984 from 1981 to 1987, then the ratio, $\overline{union_i}$ is $\frac{3}{7}$. Table 2.10 shows ($\hat{\rho}$, $\hat{\beta}$, $\hat{\alpha_0}$, $\hat{\alpha_1}$, $\hat{\alpha_2}$)=(0.3380, 0.0474, 0.8721, 0.1745, 0.0488). $\hat{\beta}$ is marginally significant and $\hat{\alpha_2}$ is not significantly different from zero. A lot of empirical literatures are raised to explore the question of union effect how equivalent workers' wage differ in union and non-union employment. While the unobserved factor that influence the sorting into union and non-union employment may also affect wage, this makes endogeneity of union variable and thus we can not just assume that the status of union is strictly exogenous. In chapter 4, I will discuss the logit model with unobserved heterogeneity to explore how the current status of union respond to the union membership in the initial period in terms of the individual workers' characteristics.

2.6 Comparison With The Other Estimators

In the section I report the results of Monte Carlo simulations which compares the conditional maximum likelihood estimator in finite sample with the GMM and conditional GLS estimators (see Blundell and Bond [1998]). I follow the notations and definitions of three GMM and CGLS estimators studied by Blundell and Bond as follows :

DIF: The standard first-differenced GMM estimator, based on moment conditions, $E(y_{i,t-s}\Delta\varepsilon_{it}) = 0$ for t=3,...,T and $s \geq 2$.

SYS: The system GMM estimator, based on linear restriction.

ALL: The system GMM estimator which also exploits the complete set of second-order moment restrictions.

CGLS: The feasible conditional GLS estimator, which uses residuals from the one-step GMM (SYS) estimator to estimate the required variance components.

I follow the data generation processes for y_{it} used by Blundell and Bond except for the y_{i0} and a_i.

$$y_{it} = \rho\, y_{i,t-1} + a_i + \varepsilon_{it}, i = 1, \ldots, N, t = 1, \ldots, T. \tag{2.50}$$

I use the same magnitude of N and T as that in Blundell and Bond paper (1998) to make the comparisons. N is chosen as 100, 200 and 500 T = 4 and 11. The true value of ρ is taken to be 0, 0.3, 0.5, 0.8, 0.9. Table 2.11 reports model (2.50) of N = 100, 200, 500 with T = 4 and Table 2.12 further reports the same model of N = 100, 200, 500 with T = 11. All results of simulation are based on 1000 Monte Carlo replications, with new values for the initial conditions drawn in each repetition.

The data generation of the first period in the model of CMLE is different from the other models in this section. The true models of GMMs and CGLS consider the generation of the initial conditions y_{i0} as :

$$y_{i0} = \frac{a_i}{1-\rho} + u_{i0}, \tag{2.51}$$

where u_{i0} is an i.i.d N(0, 4/3) random variable and independent of both a_i and ε_{it}. The variance of u_{i0} is designed to satisfy stationarity. The a_i and ε_{it} are drawn as mutually independent N(0,1) random variables. In the case of CMLE, the unobserved effects are assumed to be conditioning on the initial observations y_{i0}, so the true linear projection is assumed to be:

$$a_{i0} = 0.2 + 0.4y_{i0} + c_i, \tag{2.52}$$

where c_i is assumed to be $N(0, (\frac{1}{1-\rho})^2 + 4/3)$ and y_{i0} is generated from $N(0,1)$. The magnitude of variance of a_i in model (2.52) is designed to be equal to the variance of y_{i0} conditional on a_i in model (2.51). The contribution of the individual effects of the error terms becomes less important due to the fact that the variance increases with the ρ.

As for the non-normality assumption of errors, the comparisons among various estimators in this section will be limited on the case of $\rho = 0.5$, $\sigma_a^2 = 1$ and $N = 200$ with T=4 for the models studied by Blendell and Bond and the CMLE. Accordingly, on the one hand, the true model of GMMs and CGLS turn out to be that $y_{i0} = 2\ a_i + u_{i0}$ with $u_{i0} \sim$ N(0, 4/3), $\sigma_a^2 = 1$ while $\varepsilon_{it} = \frac{e_{it}-1}{2}$, where $e_{it} \sim \chi^2(1)$; on the other hand, the true model of CMLE generates from (2.52) in which $c_i \sim$ i.i.d N(0, 10/3) and $\varepsilon_{it} = \frac{e_{it}-1}{2}$, where $e_{it} \sim \chi^2(1)$. Table 2.13 presents a stationary design but with non-normal errors for various GMMs, GLSs and that of CMLE with non-normal errors. Table 2.14 presents the performance of the CMLE with different value of N with fixed $T = 4$ under non-normality on errors.

As is well known, when ρ is close to zero, the influence of the initial conditions becomes less important; therefore, the performance of the estimators is similar. The more interesting case is high values of ρ, which is where the GMM estimators suggested by Blundell and Bond (1998) show a clear advantage over the usual IV estimator. Table 2.11(a) shows the dramatic improvement resulting from using extra moment conditions based on restrictions of the initial conditions. for true values of ρ of 0.8 and 0.9, respectively, the Monte Carlo averages for the estimator of ρ are: $\hat{\rho}_{DIF}$ = 0.4844, 0.2264; $\hat{\rho}_{SYS-GMM}$ = 0.8101, 0.9405; $\hat{\rho}_{ALL-GMM}$ = 0.8169, 0.9422; $\hat{\rho}_{CGLS}$ = 0.8365, 0.9572; $\hat{\rho}_{CMLE}$ = 0.8004, 0.8988.

The conditional MLE has the least amount of bias, whereas the standard first-differencing IV estimator behaves very poorly. The GMM and conditional GLS estimators work better, but not as well as the CMLE. The CMLE also has the smallest standard deviations and root mean squared errors. We can summarize the findings in Table 2.11 for bias, standard deviation, and RMSE as follows:

$$|\text{Bias}(\hat{\rho}_{CMLE})| < |\text{Bias}(\hat{\rho}_{SYS-GMM})| <$$
$$|\text{Bias}(\hat{\rho}_{ALL-GMM})| < |\text{Bias}(\hat{\rho}_{CGLS})| < |\text{Bias}(\hat{\rho}_{DIF})|.$$

The ranking of corresponding standard deviations and RMSE of these estimators is as follows:

$$\text{SD}(\hat{\rho}_{CMLE}) < \text{SD}(\hat{\rho}_{ALL-GMM}) <$$
$$\text{SD}(\hat{\rho}_{SYS-GMM}) < \text{SD}(\hat{\rho}_{CGLS}) < \text{SD}(\hat{\rho}_{DIF}),$$

and

$$\text{RMSE}(\hat{\rho}_{CMLE}) < \text{RMSE}(\hat{\rho}_{ALL-GMM}) <$$
$$\text{RMSE}(\hat{\rho}_{SYS-GMM}) < \text{RMSE}(\hat{\rho}_{CGLS}) < \text{RMSE}(\hat{\rho}_{DIF}).$$

Table 2.11 shows that the performance of ALL-GMM , SYS-GMM and CMLE estimators is getting close to each other with larger N.

When T increases to 11, the bias of all estimators decreases and the standard deviations of all estimators significantly decrease. For example, from Table 2.11-

(b) and 2.12- (b), at the high value of $\rho = 0.8$, the means of $\hat{\rho}_{DIF}$ changes from 0.4844 (0.5219) to 0.7373 (0.0742); the means of $\hat{\rho}_{SYS-GMM}$ changes from 0.8050 to 0.8025; the means of $\hat{\rho}_{ALL-GMM}$ changes from 0.8112 (0.1195) to 0.8075 (0.0420); the means of $\hat{\rho}_{CGLS}$ changes from 0.8259 (0.1138) to 0.8039 (0.0423); the means of $\hat{\rho}_{CMLE}$ changes from 0.8004 (0.0684) to 0.8003 (0.0127), where the number of bracket is standard deviation.

According to the ranking of λ-class estimators of (1.27):

$$\operatorname*{plim}_{N\to\infty} \hat{\rho}(0) < \rho < \operatorname*{plim}_{N\to\infty} \hat{\rho}(\theta^2) < \operatorname*{plim}_{N\to\infty} \hat{\rho}(1) < \operatorname*{plim}_{N\to\infty} \hat{\rho}(\infty).$$

The means of estimators in Table 2.13 follow the ranking:

$$\hat{\rho}_{within}(= -0.0343) < \hat{\rho}_{GLS}(= 0.6659) < \hat{\rho}_{OLS}(= 0.8740),$$

and the estimates of the other estimates fall the range [-0.0343, 0.6659]. The comparison of Table 2.11-(b) and 2.13 suggest that the assumption of non-normality has little impact on the means and standard deviations of these estimators. At the true value of $\rho = 0.5$, the means of $\hat{\rho}_{DIF}$ changes from 0.4828 (0.1821) to 0.4867 (0.1844); the means of $\hat{\rho}_{SYS-GMM}$ changes from 0.5098 (0.0936) to 0.4999 (0.1082); the means of $\hat{\rho}_{ALL-GMM}$ changes from 0.5079 (0.0922) to 0.5067 (0.1109); the means of $\hat{\rho}_{CGLS}$ changes from 0.5135 (0.1006) to 0.5124 (0.1030); the means of $\hat{\rho}_{CMLE}$ changes from 0.5068 (0.1036) to 0.5179 (0.1227), where the number of bracket is standard deviation. Obviously, the standard deviations of all estimators become larger and the bias of all estimators enlarge a little. In Table 2.13, the standard deviations and the bias of CMLE estimator is slightly greater than GMMs and CGLS in the absence of normality assumption. Table 2.14 shows that at the true value of $\rho = 0.5$, the bias of CMLE estimator decrease almost triple and the standard deviation decrease about one and a half times to double when N increase by one time. When N is large enough the estimator of CMLE perform well in the absence of normality assumption.

2.7 Conclusion

In this chapter I consider the CMLE for the AR(1) model with unobserved effects which was proposed by Blundell and Smith (1991) in the case of no covariates. I treat the initial value in different way. Balestra and Nerlove (1966) first explored the conditional MLE, but he treat the initial value as nonrandom. It means the initial value is independent of the unobserved effects. Such assumption is usually untenable assumption. Blundell and Smith (1991) consider a range of CMLE estimators is equivalent to the ML estimator in Bhagarva and Sargan (1983), from the case without the full error components restrictions, to the fully stationary error components model. We need to care what about the restrictions on the initial value (, or distribution of $h(y_{i0}|a_i)$) and the distribution of a_i. The inclusion of x_{it} make matters even more complicate. Because we do not need impose restrictions on the y_{i0} and specify the distribution of a_i. Under the linear case the conditional ML estimators can be worked out in a simple way. The inclusion of strictly exogenous variables x_{it} will not complicates matters. This approach can be easily applied in the more complicate model, such as the state dependence model and the logit model considered in later chapters in this thesis by using the approach proposed by Wooldridge(2000b).

In practice, if we want to include the non-strictly exogenous variables, we need to specify another conditional distribution for explanatory variables X on which we do not impose strict exogeneity, $D(x_t|Y_{t-1}, Z_t, a)$ in constructing the CMLE model, where Z_t denotes the other strictly exogenous variables (see, Wooldridge [2000a]). We can let

$$D(x_t|Y_{t-1}, Z_t, a) = D(x_t|Y_{t-1}, z_t, a) \tag{2.53}$$

which means that once current z_t, past y_t and a are controlled for, z_s, $s \neq t$, has no effect on the distribution of x_t. Practically, we can parameterize the conditional density:

$$g_t(x_t|Y_{t-1}, z_t, a; \gamma_0) \tag{2.54}$$

where γ_0 is finite dimensional parameter. By equations (2.3), (2.56) and the usual product law for conditional densities, the joint density of (y_t, x_t) given $(Z_T, Y_{t-1}, X_{t-1}, a)$ is

$$p_t(w_t|W_{t-1}, z_t, a; \theta_0) = f_t(y_t|Y_{t-1}, x_t, a; \delta_0)g_t(x_t|Y_{t-1}, z_t, a; \gamma_0) \tag{2.55}$$

where $w_t = (y_t, x_t)$, $W_t = (w_t, \ldots, w_0)$ and $\theta = (\delta, \gamma)$. It follows that the density of $(w_T, \ldots, w_1)$ given (Z_T, x_0, a) is

$$p(w_T, \ldots, w_1|Z_T, w_0, a; \theta_0) = \prod_{t=1}^{T} p_t(w_t|W_{t-1}, z_t, a; \theta_0). \tag{2.56}$$

Similarly, we set up an log-likelihood function by the use of the joint conditional density function (2.56) and conditional density function for a, similar to function (2.5) to integrate out the unobserved effects. The question can be written as follows:

$$\log \int_{\mathbb{R}} \prod_{t=1}^{T} p_t(w_{it}|w_{i,t-1}, z_{it}, a; \theta)h(a|w_{i0}, z_i; \lambda)\, v(\mathrm{d}\, a).$$

If we have random sampling in the cross section dimension and standard regularity conditions, with fixed T the CMLE for ϑ_0 will be consistent and $\sqrt{N}$-asymptotically normally distributed. (See Newey and McFadden [1994] for sufficient regularity conditions.) But it will be computationally difficult, especially in the wage-union application: union would have to follow a dynamic probit or logit model, as in Chapter 4.

In the previous simulation, I employ the Hermite integral formula, $\int_{-\infty}^{\infty} f(z)e^{-z^2}\, \mathrm{d}\, x \simeq \sum_{j=1}^{k} f(z_j)w_j$, but this computation is costly. When we need to include the conditional density of non-strictly exogenous variables in the integration, the problem of calculating the integration grow burdensome. In the model, although we do not need full distributional assumption on the non-strictly exogenous variables and the unobserved effects for consistent estimation, we need measure how sensitive are the estimates of important quantities to the specifications of (2.5), e.g. the α_0, α_1, or α_2.

Table 2.1: $H_0: \theta = \theta_0$, where ρ =0 ~ 0.95

$$\theta = (\rho, 0.2, 0.4, \sqrt{2.4}, \sqrt{1.2})$$
$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{2.4}, \sqrt{1.2})$$

P\\$\hat{\theta}$	5×10^{-4}	0.1990	0.3996	1.5482	1.0841	ρ_0
0.01	0.0108	0.0075	0.0092	0.0150	0.0117	
0.05	0.0692	0.0442	0.0458	0.0542	0.0558	0
0.10	0.1142	0.0933	0.0967	0.0958	0.1175	
P\\$\hat{\theta}$	0.1012	0.2007	0.4012	1.5481	1.0840	ρ_0
0.01	0.0133	0.0083	0.0092	0.0158	0.0108	
0.05	0.0692	0.0442	0.0467	0.0533	0.0567	0.1
0.10	0.1008	0.0925	0.0950	0.1000	0.1550	
$P\backslash\hat{\theta}$	0.1512	0.2006	0.4012	1.5481	1.0839	ρ_0
0.01	0.0125	0.0083	0.0100	0.0158	0.0100	
0.05	0.0683	0.0442	0.0483	0.0500	0.0550	0.15
0.10	0.1125	0.0925	0.1000	0.1025	0.1133	
$P\backslash\hat{\theta}$	0.2013	0.2006	0.4011	1.5482	1.0837	ρ_0
0.01	0.0117	0.0083	0.0117	0.0158	0.0092	
0.05	0.0650	0.0442	0.0492	0.0500	0.0550	0.2
0.10	0.1108	0.0933	0.1025	0.1008	0.1117	

Normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200} \theta_j^*$, $\sqrt{2.4} \simeq 1.5492$, $\sqrt{1.2} \simeq 1.0954$

Continue (a)

$$\theta = (\rho, 0.2, 0.4, \sqrt{2.4}, \sqrt{1.2})$$

$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{2.4}, \sqrt{1.2})$$

P\\$\hat{\theta}$	0.2517	0.2006	0.4011	1.5483	1.0835	ρ_0
0.01	0.0125	0.0075	0.0117	0.0158	0.0083	
0.05	0.0625	0.0442	0.0483	0.0508	0.0558	0.25
0.10	0.1125	0.0933	0.1033	0.1025	0.1125	
P\\$\hat{\theta}$	0.3014	0.2006	0.4010	1.5484	1.0833	ρ_0
0.01	0.0125	0.0075	0.0108	0.0158	0.0083	
0.05	0.0625	0.0442	0.0483	0.0517	0.0550	0.3
0.10	0.1108	0.0942	0.1000	0.0992	0.1125	
$P\backslash\hat{\theta}$	0.3514	0.2006	0.4009	1.5484	1.0830	ρ_0
0.01	0.0117	0.0075	0.0100	0.0150	0.0083	
0.05	0.0583	0.0433	0.0492	0.0508	0.0508	0.35
0.10	0.1092	0.0933	0.1025	0.0975	0.1108	
$P\backslash\hat{\theta}$	0.3944	0.2022	0.4035	1.5473	1.0888	ρ_0
0.01	0.0142	0.0067	0.0067	0.0150	0.0100	
0.05	0.0508	0.0400	0.0500	0.0508	0.0492	0.4
0.10	0.1117	0.0858	0.1000	0.0992	0.1092	
$P\backslash\hat{\theta}$	0.4517	0.2005	0.4006	1.5487	1.0822	ρ_0
0.01	0.0125	0.0075	0.0092	0.0142	0.0092	
0.05	0.0525	0.0433	0.0542	0.0550	0.0500	0.45
0.10	0.1083	0.0925	0.1017	0.0925	0.1075	

Normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{2.4} \simeq 1.5492$, $\sqrt{1.2} \simeq 1.0954$

Continue (b)

$$\theta = (\rho, 0.2, 0.4, \sqrt{2.4}, \sqrt{1.2})$$

$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{2.4}, \sqrt{1.2})$$

$P\backslash\hat{\theta}$	0.5021	0.2008	0.3998	1.5491	1.0811	ρ_0
0.01	0.0100	0.0075	0.0108	0.0150	0.0100	
0.05	0.0500	0.0425	0.0525	0.0525	0.0467	0.5
0.10	0.1050	0.0942	0.1071	0.0867	0.1017	
$P\backslash\hat{\theta}$	0.5520	0.2005	0.4001	1.5490	1.0812	ρ_0
0.01	0.0083	0.0092	0.0108	0.0150	0.0058	
0.05	0.0467	0.0417	0.0525	0.0525	0.0442	0.55
0.10	0.1058	0.0967	0.0950	0.0842	0.0933	
$P\backslash\hat{\theta}$	0.6022	0.2004	0.3998	1.5492	1.0806	ρ_o
0.01	0.0083	0.0092	0.0117	0.0158	0.0050	
0.05	0.0458	0.0408	0.0542	0.0533	0.0392	0.6
0.10	0.1017	0.0950	0.0958	0.0850	0.0925	
$P\backslash\hat{\theta}$	0.6517	0.2012	0.4008	1.5490	1.0811	ρ_0
0.01	0.0075	0.0092	0.0108	0.0150	0.0050	
0.05	0.0475	0.0400	0.0558	0.0508	0.0350	0.65
0.10	0.0958	0.0900	0.1042	0.0825	0.0883	
$P\backslash\hat{\theta}$	0.7023	0.2004	0.3996	1.5493	1.0801	ρ_0
0.01	0.0100	0.0092	0.0108	0.0158	0.0042	
0.05	0.0500	0.0425	0.0542	0.0517	0.0317	0.7
0.10	0.0933	0.0942	0.0950	0.0858	0.0833	

Normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{2.4} \simeq 1.5492$, $\sqrt{1.2} \simeq 1.0954$

Continue (c)

$$\theta = (\rho, 0.2, 0.4, \sqrt{2.4}, \sqrt{1.2})$$

$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{2.4}, \sqrt{1.2})$$

P\\$\hat{\theta}$	0.7525	0.2003	0.3993	1.5495	1.0791	ρ_0
0.01	0.0108	0.0067	0.0108	0.0083	0.0067	
0.05	0.0383	0.0350	0.0425	0.0383	0.0267	0.75
0.10	0.0750	0.0714	0.0742	0.0708	0.0633	
P\\$\hat{\theta}$	0.8019	0.2005	0.3999	1.5491	1.0813	ρ_0
0.01	0.0100	0.0100	0.0083	0.0117	0.0042	
0.05	0.0467	0.0433	0.0542	0.0492	0.0383	0.8
0.10	0.0983	0.0983	0.0892	0.0875	0.0808	
$P\backslash\hat{\theta}$	0.8521	0.2004	0.3995	1.5493	1.0802	ρ_o
0.01	0.0108	0.0100	0.0100	0.0108	0.0050	
0.05	0.0442	0.0425	0.0508	0.0508	0.0367	0.85
0.10	0.0958	0.0975	0.0917	0.0942	0.0858	
$P\backslash\hat{\theta}$	0.9018	0.2005	0.3998	1.5491	1.0809	ρ_0
0.01	0.0100	0.0100	0.0108	0.0108	0.0050	
0.05	0.0458	0.0442	0.0475	0.0525	0.0383	0.9
0.10	0.0967	0.0958	0.0933	0.0900	0.0875	
$P\backslash\hat{\theta}$	0.9515	0.2005	0.4001	1.5489	1.0816	ρ_0
0.01	0.0108	0.0092	0.0108	0.0108	0.0067	
0.05	0.0467	0.0433	0.0483	0.0517	0.0433	0.95
0.10	0.0983	0.0933	0.0967	0.0892	0.0900	

Normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{2.4} \simeq 1.5492$, $\sqrt{1.2} \simeq 1.0954$

(d)

Table 2.2: $H_0 : \theta = \theta_0$, where ρ =0 $\sim$ 0.95

$$\begin{aligned}\theta &= (\rho, 0.2, 0.4, \sqrt{2.4}, \sqrt{1.2}) \\ \theta_0 &= (\rho_0, 0.2, 0.4, \sqrt{2.4}, \sqrt{1.2})\end{aligned}$$

P\ $\rho_0 \overset{=}{\to}$	0	0.05	0.1	0.15	0.2	0.25	0.3	ρ
0.01	0.0108	0.1167	0.5400	0.9050	0.9925	1.0000	1.0000	
0.05	0.0692	0.2883	0.7525	0.9608	0.9983	1.0000	1.0000	0
0.10	0.1142	0.3825	0.8367	0.9817	0.9992	1.0000	1.0000	
P\ $\rho_0 \overset{=}{\to}$	0	0.05	0.1	0.15	0.2	0.25	0.3	ρ
0.01	0.5100	0.1000	0.0013	0.1108	0.5000	0.8817	0.9892	
0.05	0.7450	0.2575	0.0692	0.2667	0.7117	0.9525	0.9975	0.1
0.10	0.8375	0.3767	0.1108	0.3758	0.8083	0.9717	0.9992	
P\ $\rho_0 \overset{=}{\to}$	0	0.05	0.1	0.15	0.2	0.25	0.3	ρ
0.01	0.9100	0.4950	0.0950	0.0125	0.1100	0.4842	0.8617	
0.05	0.9708	0.7250	0.2525	0.0683	0.2542	0.6967	0.9433	0.15
0.10	0.9883	0.8267	0.3633	0.1125	0.3667	0.7975	0.9667	
P\ $\rho_0 \overset{=}{\to}$	0	0.05	0.1	0.15	0.2	0.3	0.35	ρ
0.01	0.9950	0.8967	0.4742	0.0900	0.0117	0.4700	0.8467	
0.05	1.0000	0.9692	0.7117	0.2425	0.0650	0.6850	0.9400	0.2
0.10	1.0000	0.9842	0.8133	0.3542	0.1108	0.7825	1.0000	

Normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200} \theta_j^*$, $\sqrt{2.4} \simeq 1.5492$, $\sqrt{1.2} \simeq 1.0954$

Continue (a)

$$\theta = (\rho, 0.2, 0.4, \sqrt{2.4}, \sqrt{1.2})$$

$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{2.4}, \sqrt{1.2})$$

P\ $\rho_0 \overset{=}{\rightarrow}$	0.1	0.15	0.2	0.3	0.35	0.4	0.45	ρ
0.01	0.8833	0.4600	0.0842	0.1033	0.4542	0.8317	0.9708	
0.05	0.9675	0.6950	0.2283	0.2542	0.6675	0.9325	0.9942	0.25
0.10	0.9783	0.8000	0.3417	0.3533	0.7650	0.9542	0.9967	
P\ $\rho_0 \overset{=}{\rightarrow}$	0.15	0.2	0.25	0.35	0.4	0.45	0.5	ρ
0.01	0.8725	0.4358	0.0800	0.1042	0.4325	0.8125	0.9650	
0.05	0.9625	0.6767	0.2192	0.2475	0.6450	0.9233	0.9933	0.3
0.10	0.9775	0.7817	0.3358	0.3400	0.7542	0.9500	0.9967	
P\ $\rho_0 \overset{=}{\rightarrow}$	0.2	0.25	0.3	0.4	0.45	0.5	0.55	ρ
0.01	0.8600	0.4117	0.0758	0.1025	0.4133	0.7933	0.9558	
0.05	0.9567	0.6633	0.2083	0.2383	0.6358	0.9092	0.9900	0.35
0.10	0.9750	0.7750	0.3208	0.3325	0.7408	0.9442	0.9933	
P\ $\rho_0 \overset{=}{\rightarrow}$	0.25	0.3	0.35	0.45	0.5	0.55	0.6	ρ
0.01	0.8333	0.3733	0.0625	0.1083	0.4183	0.7808	0.9533	
0.05	0.9500	0.6333	0.1933	0.2458	0.6425	0.9050	0.9950	0.4
0.10	0.9742	0.7425	0.2942	0.3417	0.7425	0.9450	0.9967	
P\ $\rho_0 \overset{=}{\rightarrow}$	0.3	0.35	0.4	0.5	0.55	0.6	0.65	ρ
0.01	0.8333	0.3742	0.0625	0.0983	0.3858	0.7542	0.9383	
0.05	0.9492	0.6450	0.1983	0.2292	0.6175	0.8825	0.9808	0.45
0.10	0.9742	0.7475	0.2983	0.3150	0.7192	0.9292	0.9917	

Normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{2.4} \simeq 1.5492$, $\sqrt{1.2} \simeq 1.0954$

Continue (b)

$$\theta = (\rho, 0.2, 0.4, \sqrt{2.4}, \sqrt{1.2})$$

$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{2.4}, \sqrt{1.2})$$

P \ $\rho_0 \overset{=}{\rightarrow}$	0.35	0.4	0.45	0.55	0.6	0.65	0.7	ρ
0.01	0.8200	0.3525	0.0575	0.7442	0.9317	0.9883	0.9975	
0.05	0.9450	0.6225	0.1842	0.8667	0.9767	0.9950	0.9992	0.5
0.10	0.9733	0.7442	0.2992	0.9242	0.9900	0.9983	1.0000	
P \ $\rho_0 \overset{=}{\rightarrow}$	0.4	0.45	0.5	0.6	0.65	0.7	0.75	ρ
0.01	0.8183	0.3442	0.0517	0.1075	0.3783	0.7317	0.9225	
0.05	0.9433	0.6192	0.1800	0.2258	0.6033	0.8625	0.9733	0.55
0.10	0.9742	0.7425	0.2933	0.3158	0.6992	0.9133	0.9867	
P \ $\rho_0 \overset{=}{\rightarrow}$	0.45	0.5	0.55	0.65	0.7	0.75	0.8	ρ
0.01	0.8217	0.3408	0.0392	0.1067	0.3792	0.7208	0.9117	
0.05	0.9425	0.6200	0.1758	0.2308	0.6008	0.8600	0.9675	0.6
0.10	0.9742	0.7433	0.2850	0.3125	0.6875	0.9058	0.9850	
P \ $\rho_0 \overset{=}{\rightarrow}$	0.5	0.55	0.6	0.7	0.75	0.8	0.85	ρ
0.01	0.8275	0.3417	0.0308	0.1117	0.3958	0.7283	0.9175	
0.05	0.9450	0.6233	0.1758	0.2425	0.5933	0.8583	0.9667	0.65
0.10	0.9775	0.7383	0.2808	0.3300	0.6842	0.9075	0.9842	
P \ $\rho_0 \overset{=}{\rightarrow}$	0.55	0.6	0.65	0.75	0.8	0.85	0.9	ρ
0.01	0.8558	0.3633	0.0292	0.1192	0.4108	0.7350	0.9800	
0.05	0.9550	0.6442	0.1758	0.2425	0.5983	0.8742	0.9917	0.7
0.10	0.9833	0.7725	0.2958	0.3292	0.6908	0.9150	0.9975	

Normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{2.4} \simeq 1.5492$, $\sqrt{1.2} \simeq 1.0954$

Continue (c)

$$\theta = (\rho, 0.2, 0.4, \sqrt{2.4}, \sqrt{1.2})$$

$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{2.4}, \sqrt{1.2})$$

P \ $\rho_0 \rightarrow$	0.6	0.65	0.7	0.8	0.85	0.9	0.95	ρ
0.01	0.7142	0.3317	0.0283	0.0975	0.3458	0.6183	0.7533	
0.05	0.7867	0.5617	0.1567	0.2000	0.5067	0.7183	0.7900	0.75
0.10	0.7967	0.6500	0.2550	0.2750	0.5767	0.7517	0.7967	
P \ $\rho_0 =$ $\rightarrow$	0.65	0.7	0.75	0.85	0.9	0.95	1	ρ
0.01	0.9300	0.4592	0.0442	0.1292	0.8058	0.9483	0.9908	
0.05	0.9842	0.7442	0.2108	0.2767	0.9092	0.9817	0.9967	0.8
0.10	0.9942	0.8358	0.3350	0.3725	0.9392	0.9917	0.9992	
P \ $\rho_0 =$ $\rightarrow$	0.65	0.7	0.75	0.8	0.9	0.95	1	ρ
0.01	1.0000	0.9617	0.5392	0.0625	0.1458	0.5333	0.8475	
0.05	1.0000	0.9925	0.7958	0.2475	0.2942	0.7125	0.9342	0.85
0.10	1.0000	0.9975	0.8792	0.3650	0.3992	0.7958	0.9600	
P \ $\rho_0 =$ $\rightarrow$	0.65	0.7	0.75	0.8	0.85	0.95	1	ρ
0.01	1.0000	1.0000	0.9850	0.6558	0.0908	0.1658	0.5975	
0.05	1.0000	1.0000	0.9975	0.8542	0.2858	0.3258	0.7717	0.9
0.10	1.0000	1.0000	1.0000	0.9258	0.4067	0.4392	0.8358	
P \ $\rho_0 =$ $\rightarrow$	0.7	0.75	0.8	0.85	0.9	0.95	1	ρ
0.01	1.0000	1.0000	0.9942	0.7583	0.1258	0.0108	0.1933	
0.05	1.0000	1.0000	1.0000	0.9208	0.3392	0.0467	0.3725	0.95
0.10	1.0000	1.0000	1.0000	0.9208	0.3392	0.0467	0.3725	

Normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{2.4} \simeq 1.5492$, $\sqrt{1.2} \simeq 1.0954$

(d)

Table 2.3: $H_0 : \theta = \theta_0$, where ρ =0 $\sim$ 0.95

$$\theta \;= (\rho, 0.2, 0.4, \sqrt{1.5}, \sqrt{1.25})$$
$$\theta_0 \;= (\rho_0, 0.2, 0.4, \sqrt{1.5}, \sqrt{1.25})$$

$P\backslash\hat{\theta}$	8×10^{-5}	0.2015	0.3982	1.2226	1.1095	ρ_0
0.01	0.0092	0.0092	0.0133	0.0725	0.0200	
0.05	0.0467	0.0575	0.0600	0.1442	0.0783	0
0.10	0.0858	0.0958	0.1108	0.2242	0.1433	
P$\backslash\hat{\theta}$	0.0999	0.2044	0.3971	1.2225	1.1090	ρ_0
0.01	0.0108	0.0100	0.0108	0.0700	0.0192	
0.05	0.0467	0.0550	0.0600	0.1417	0.0742	0.1
0.10	0.0892	0.0942	0.1125	0.2192	0.1425	
P$\backslash\hat{\theta}$	0.1499	0.2044	0.3970	1.2225	1.1090	ρ_0
0.01	0.0108	0.0100	0.0108	0.0683	0.0192	
0.05	0.0475	0.0558	0.0592	0.1442	0.0742	0.15
0.10	0.0858	0.0958	0.1092	0.2208	0.1433	
P$\backslash\hat{\theta}$	0.2000	0.2044	0.3970	1.2226	1.1090	ρ_0
0.01	0.0100	0.0108	0.0117	0.0667	0.0183	
0.05	0.0475	0.0550	0.0608	0.1442	0.0742	0.2
0.10	0.0850	0.0967	0.1092	0.2208	0.1425	

Non-normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{1.5} \simeq 1.2247$,$\sqrt{1.25} \simeq 1.1180$

Continue (a)

$$\theta = (\rho, 0.2, 0.4, \sqrt{1.5}, \sqrt{1.25})$$

$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{1.5}, \sqrt{1.25})$$

P\\$\hat{\theta}$	0.2500	0.2044	0.3970	1.2226	1.1090	ρ_0
0.01	0.0117	0.0108	0.0117	0.0642	0.0158	
0.05	0.0467	0.0550	0.0600	0.1450	0.0742	0.25
0.10	0.0833	0.0958	0.1050	0.2225	0.1392	
P\\$\hat{\theta}$	0.3000	0.2044	0.3970	1.2226	1.1090	ρ_0
0.01	0.0108	0.0108	0.0117	0.0625	0.0158	
0.05	0.0467	0.0542	0.0600	0.1442	0.0700	0.3
0.10	0.0867	0.0958	0.1075	0.2217	0.1317	
$P\backslash\hat{\theta}$	0.3500	0.2044	0.3970	1.2226	1.1090	ρ_o
0.01	0.0125	0.0100	0.0117	0.0558	0.0167	
0.05	0.0508	0.0533	0.0600	0.1450	0.0675	0.35
0.10	0.0892	0.0950	0.1050	0.2192	0.1250	
$P\backslash\hat{\theta}$	0.4000	0.2044	0.3970	1.2226	1.1091	ρ_0
0.01	0.0125	0.0125	0.0117	0.0558	0.0175	
0.05	0.0525	0.0533	0.0575	0.1442	0.0683	0.4
0.10	0.0950	0.0967	0.1033	0.2150	0.1233	
$P\backslash\hat{\theta}$	0.4500	0.2044	0.3970	1.2226	1.1091	ρ_0
0.01	0.0117	0.0083	0.0125	0.0525	0.0183	
0.05	0.0508	0.0517	0.0592	0.1417	0.0658	0.45
0.10	0.0983	0.0958	0.1067	0.2150	0.1217	

Non-normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{1.5} \simeq 1.2247$, $\sqrt{1.25} \simeq 1.1180$

Continue (b)

$$\theta = (\rho, 0.2, 0.4, \sqrt{1.5}, \sqrt{1.25})$$

$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{1.5}, \sqrt{1.25})$$

P\\$\hat{\theta}$	0.4999	0.2045	0.3970	1.2226	1.1093	ρ_0
0.01	0.0100	0.0100	0.0133	0.0550	0.0175	
0.05	0.0558	0.0517	0.0608	0.1400	0.0658	0.5
0.10	0.1017	0.0967	0.1033	0.2133	0.1267	
P\\$\hat{\theta}$	0.5499	0.2045	0.3970	1.2226	1.1094	ρ_0
0.01	0.0125	0.0092	0.0125	0.0550	0.0167	
0.05	0.0567	0.0525	0.0575	0.1367	0.0633	0.55
0.10	0.1017	0.0967	0.1075	0.2125	0.1308	
P\\$\hat{\theta}$	0.5999	0.2045	0.3971	1.2226	1.1095	ρ_0
0.01	0.0125	0.0100	0.0125	0.0533	0.0150	
0.05	0.0575	0.0525	0.0575	0.1392	0.0608	0.6
0.10	0.1017	0.0967	0.1092	0.2150	0.1283	
P\\$\hat{\theta}$	0.6498	0.2046	0.3972	1.2225	1.1097	ρ_0
0.01	0.0142	0.0092	0.0117	0.0542	0.0150	
0.05	0.0575	0.0533	0.0558	0.1408	0.0625	0.65
0.10	0.1017	0.0967	0.1100	0.2183	0.1242	
P\\$\hat{\theta}$	0.6998	0.2046	0.3973	1.2225	1.1099	ρ_0
0.01	0.0150	0.0092	0.0133	0.0533	0.0142	
0.05	0.0575	0.0542	0.0550	0.1400	0.0642	0.7
0.10	0.1025	0.0967	0.1100	0.2208	0.1242	

Non-normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{1.5} \simeq 1.2247$,$\sqrt{1.25} \simeq 1.1180$

Continue (c)

$$\theta = (\rho, 0.2, 0.4, \sqrt{1.5}, \sqrt{1.25})$$

$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{1.5}, \sqrt{1.25})$$

$P\backslash\hat{\theta}$	0.7498	0.2047	0.3974	1.2225	1.1101	ρ_0
0.01	0.0142	0.0092	0.0133	0.0517	0.0142	
0.05	0.0600	0.0542	0.0542	0.1375	0.0692	0.75
0.10	0.1000	0.0958	0.1083	0.2200	0.1200	

$P\backslash\hat{\theta}$	0.7997	0.2047	0.3976	1.2224	1.1103	ρ_0
0.01	0.0158	0.0100	0.0117	0.0508	0.0117	
0.05	0.0617	0.0542	0.0550	0.1433	0.0683	0.8
0.10	0.1033	0.0967	0.1108	0.2258	0.1183	

$P\backslash\hat{\theta}$	0.8497	0.2048	0.3977	1.2223	1.1105	ρ_0
0.01	0.0175	0.0092	0.0125	0.0492	0.0133	
0.05	0.0617	0.0542	0.0542	0.1408	0.0675	0.85
0.10	0.0975	0.0967	0.1125	0.2258	0.1208	

$P\backslash\hat{\theta}$	0.8997	0.2048	0.3979	1.2223	1.1107	ρ_0
0.01	0.0175	0.0108	0.0117	0.0508	0.0142	
0.05	0.0617	0.0542	0.0558	0.1142	0.0650	0.9
0.10	0.1025	0.0983	0.1175	0.2225	0.1242	

$P\backslash\hat{\theta}$	0.9496	0.2048	0.3980	1.2222	1.1108	ρ_0
0.01	0.0175	0.0100	0.0117	0.0525	0.0167	
0.05	0.0583	0.0533	0.0550	0.1475	0.0658	0.95
0.10	0.1058	0.1000	0.1200	0.2242	0.1258	

Non-normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{1.5} \simeq 1.2247$, $\sqrt{1.25} \simeq 1.1180$

(d)

Table 2.4: $H_0 : \theta = \theta_0$, where ρ =0 $\sim$ 0.95

$$\theta = (\rho, 0.2, 0.4, \sqrt{1.5}, \sqrt{1.25})$$
$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{1.5}, \sqrt{1.25})$$

P\ $\rho_0 \overset{=}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	ρ
0.01	0.0092	0.1258	0.5900	0.9433	0.9992	1.0000	1.0000	1.0000	
0.05	0.0467	0.3133	0.7808	0.9850	0.9992	1.0000	1.0000	1.0000	0
0.10	0.0858	0.4158	0.8733	0.9958	1.0000	1.0000	1.0000	1.0000	
P \ $\rho_0 \overset{=}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	ρ
0.01	0.5592	0.0933	0.0108	0.1267	0.5583	0.9267	0.9975	1.0000	
0.05	0.7833	0.2725	0.0467	0.2967	0.7633	0.9758	0.9992	1.0000	0.1
0.10	0.8800	0.3800	0.0892	0.3992	0.8583	0.9900	0.9992	1.0000	
P \ $\rho_0 \overset{=}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	ρ
0.01	0.9533	0.5442	0.0883	0.0108	0.1283	0.5500	0.9183	0.9967	
0.05	0.9875	0.7717	0.2650	0.0475	0.2942	0.7575	0.9742	0.9992	0.15
0.10	0.9917	0.8700	0.3758	0.0858	0.3950	0.8483	0.9875	0.9992	
P \ $\rho_0 \overset{=}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	ρ
0.01	0.9983	0.9492	0.5333	0.0825	0.0100	0.1250	0.5408	0.9108	
0.05	1.0000	0.9867	0.7658	0.2642	0.0475	0.2875	0.7492	0.9683	0.2
0.10	1.0000	0.9908	0.8617	0.3650	0.0850	0.3925	0.8433	0.9833	

Non-normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{1.5} \simeq 1.2247$,$\sqrt{1.25} \simeq 1.1180$

Continue (a)

$$\theta = (\rho, 0.2, 0.4, \sqrt{1.5}, \sqrt{1.25})$$

$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{1.5}, \sqrt{1.25})$$

P \ $\rho_0 \overset{=}{\rightarrow}$	0.1	0.15	0.2	0.25	0.3	0.35	0.45	0.5	ρ
0.01	0.9408	0.5183	0.0817	0.0117	0.1233	0.5342	0.9058	0.9900	
0.05	0.9867	0.7525	0.2517	0.0467	0.2825	0.7450	0.9675	0.9675	0.25
0.10	0.9917	0.8583	0.3658	0.0833	0.3858	0.8392	0.9808	0.9992	
P \ $\rho_0 \overset{=}{\rightarrow}$	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	ρ
0.01	0.9375	0.5100	0.0800	0.0108	0.1250	0.5242	0.8992	0.9858	
0.05	0.9858	0.7467	0.2425	0.0467	0.2800	0.7375	0.9658	0.9983	0.3
0.10	0.9908	0.8467	0.3600	0.0867	0.3825	0.8367	0.9775	0.9992	
P \ $\rho_0 \overset{=}{\rightarrow}$	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	ρ
0.01	0.9342	0.5083	0.0775	0.0125	0.1208	0.5208	0.8950	0.9850	
0.05	0.9867	0.7442	0.2367	0.0508	0.2792	0.7375	0.9650	0.9983	0.35
0.10	0.9917	0.8458	0.3650	0.0892	0.3833	0.8350	0.9775	0.9992	
P \ $\rho_0 \overset{=}{\rightarrow}$	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	ρ
0.01	0.9317	0.5025	0.0758	0.0125	0.1167	0.5133	0.8908	0.9825	
0.05	0.9850	0.7400	0.2317	0.0525	0.2825	0.7375	0.9617	0.9975	0.4
0.10	0.9925	0.8467	0.3633	0.0950	0.3850	0.8292	0.9758	0.9983	
P \ $\rho_0 \overset{=}{\rightarrow}$	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	ρ
0.01	0.9383	0.5058	0.0767	0.0117	0.1142	0.5167	0.8833	0.9825	
0.05	0.9875	0.7442	0.2292	0.0508	0.2858	0.7425	0.9583	0.9967	0.45
0.10	0.9925	0.8517	0.3642	0.0983	0.3908	0.8258	0.9758	0.9975	

Non-normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{1.5} \simeq 1.2247$, $\sqrt{1.25} \simeq 1.1180$

Continue (b)

$$\theta = (\rho, 0.2, 0.4, \sqrt{1.5}, \sqrt{1.25})$$

$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{1.5}, \sqrt{1.25})$$

P \ $\rho_0 \overset{=}{\to}$	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	ρ
0.01	0.9392	0.5125	0.0792	0.0100	0.1183	0.5217	0.8842	0.9842	
0.05	0.9875	0.7542	0.2317	0.0558	0.2950	0.7450	0.9575	0.9950	0.5
0.10	0.9950	0.8558	0.3667	0.1017	0.3900	0.8308	0.9767	0.9975	
P \ $\rho_0 \overset{=}{\to}$	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	ρ
0.01	0.9467	0.5350	0.0825	0.0125	0.1292	0.5408	0.8950	0.9842	
0.05	0.9883	0.7700	0.2383	0.0567	0.3083	0.7500	0.9558	0.9950	0.55
0.10	0.9950	0.8675	0.3617	0.1017	0.3933	0.8375	0.9792	0.9975	
P \ $\rho_0 \overset{=}{\to}$	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	ρ
0.01	0.9550	0.5583	0.0867	0.0125	0.1383	0.5683	0.9042	0.9850	
0.05	0.9908	0.7917	0.2475	0.0575	0.3142	0.7592	0.9583	0.9958	0.6
0.10	0.9950	0.8792	0.3700	0.1017	0.4150	0.8425	0.9800	0.9967	
P \ $\rho_0 \overset{=}{\to}$	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	ρ
0.01	0.9642	0.5917	0.1008	0.0142	0.1575	0.6000	0.9117	0.9892	
0.05	0.9908	0.8167	0.2617	0.0575	0.3308	0.7892	0.9650	0.9958	0.65
0.10	0.9958	0.8992	0.3482	0.1017	0.4267	0.8517	0.9817	0.9975	
P \ $\rho_0 \overset{=}{\to}$	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	ρ
0.01	0.9767	0.6425	0.1150	0.0150	0.1700	0.6400	0.9242	0.9917	
0.05	0.9933	0.8508	0.2858	0.0575	0.3592	0.8067	0.9758	0.9975	0.7
0.10	0.9967	0.9175	0.4008	0.1025	0.4550	0.8650	0.9867	0.9975	

Non-normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{1.5} \simeq 1.2247$, $\sqrt{1.25} \simeq 1.1180$

Continue (c)

$$\theta = (\rho, 0.2, 0.4, \sqrt{1.5}, \sqrt{1.25})$$

$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{1.5}, \sqrt{1.25})$$

P \ $\rho_0 \overset{=}{\rightarrow}$	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	ρ
0.01	0.9858	0.7042	0.1308	0.0142	0.1850	0.6750	0.9408	0.9975	
0.05	0.9950	0.8867	0.3083	0.0600	0.3767	0.8258	0.9833	1.0000	0.75
0.10	0.9975	0.9325	0.4283	0.1000	0.4842	0.8883	0.9892	1.0000	
P \ $\rho_0 \overset{=}{\rightarrow}$	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	ρ
0.01	0.9900	0.7725	0.1533	0.0158	0.2175	0.7250	0.9608	0.9967	
0.05	0.9975	0.9192	0.3425	0.0617	0.4042	0.8642	0.9892	0.9975	0.8
0.10	0.9992	0.9558	0.4708	0.1033	0.5208	0.9108	0.9942	0.9992	
P \ $\rho_0 \overset{=}{\rightarrow}$	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	ρ
0.01	1.0000	0.9933	0.8375	0.1792	0.0175	0.2492	0.7750	0.9775	
0.05	1.0000	0.9992	0.9417	0.3908	0.0617	0.4608	0.9017	0.9942	0.85
0.10	1.0000	0.9992	0.9642	0.5200	0.0975	0.5625	0.9375	0.9967	
P \ $\rho_0 \overset{=}{\rightarrow}$	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	ρ
0.01	1.0000	1.0000	0.9992	0.8950	0.2208	0.0175	0.2908	0.8367	
0.05	1.0000	1.0000	0.9992	0.9650	0.4442	0.0617	0.4933	0.9333	0.9
0.10	1.0000	1.0000	0.9992	0.9867	0.5833	0.1025	0.6000	0.9625	
P \ $\rho_0 \overset{=}{\rightarrow}$	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	ρ
0.01	1.0000	1.0000	1.0000	0.9992	0.9367	0.2833	0.0175	0.3425	
0.05	1.0000	1.0000	1.0000	0.9992	0.9858	0.5200	0.0583	0.5600	0.95
0.10	1.0000	1.0000	1.0000	1.0000	0.9900	0.6533	0.1058	0.6517	

Non-normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200} \sum_{j=1}^{1200} \theta_j^*$, $\sqrt{1.5} \simeq 1.2247$, $\sqrt{1.25} \simeq 1.1180$

(d)

Table 2.5: CMLE for the dynamic panel data of log-wage with unobserved heterogeneity, period:1980 $\sim$ 1987

coefficient	$\hat{\rho}$	$\hat{\alpha_0}$	$\hat{\alpha_1}$	$\hat{\sigma}_\varepsilon$	$\hat{\sigma}_a$
CMLE	0.3405	0.8784	0.1839	0.3511	0.2162
t-statistics	(18.418)	(25.328)	(8.704)	(77.425)	(18.904)

Table 2.6: $H_0 : \theta = \theta_0,\ where\ \rho = 0 \sim 0.95$

$$\begin{cases} \theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{2.4}, \sqrt{1.2}) \\ \theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{2.4}, \sqrt{1.2}) \end{cases}$$

P\\$\hat{\theta}$	1.7×10^{-3}	0.1500	0.1988	0.4029	0.3528	1.5478	1.0801	ρ_0
0.01	0.0100	0.0083	0.0100	0.0100	0.0142	0.0108	0.0142	
0.05	0.0583	0.0517	0.0625	0.0492	0.0483	0.0533	0.0483	0
0.10	0.1058	0.1033	0.1067	0.1025	0.0917	0.1033	0.1067	
P\\$\hat{\theta}$	0.1017	0.1500	0.1988	0.4028	0.3527	1.5479	1.0798	ρ_0
0.01	0.0108	0.0083	0.0100	0.0100	0.0133	0.0108	0.0142	
0.05	0.0567	0.0517	0.0617	0.0457	0.0508	0.0517	0.0500	0.1
0.10	0.1033	0.1033	0.1058	0.1033	0.0908	0.1058	0.0142	
P\\$\hat{\theta}$	0.1571	0.1500	0.1988	0.4027	0.3527	1.5479	1.0797	ρ_0
0.01	0.0100	0.0083	0.0108	0.0108	0.0133	0.0108	0.0142	
0.05	0.0583	0.0517	0.0617	0.0483	0.0500	0.0508	0.0492	0.15
0.10	0.1025	0.1033	0.1058	0.1017	0.0900	0.1042	0.0142	
P\\$\hat{\theta}$	0.2017	0.1500	0.1988	0.4027	0.3527	1.5479	1.0795	ρ_0
0.01	0.0117	0.0083	0.0108	0.0108	0.0133	0.0117	0.0142	
0.05	0.0567	0.0517	0.0617	0.0508	0.0492	0.0500	0.0492	0.2
0.10	0.1017	0.1025	0.1042	0.1000	0.0917	0.1050	0.1083	

Normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{2.4} \simeq 1.5942, \sqrt{1.2} \simeq 1.0954$

Continue (a)

$$\begin{cases} \theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{2.4}, \sqrt{1.2}) \\ \theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{2.4}, \sqrt{1.2}) \end{cases}$$

P\\$\hat{\theta}$	0.2517	0.1500	0.1988	0.4026	0.3526	1.5480	1.0794	ρ_0
0.01	0.0125	0.0083	0.0108	0.0108	0.0133	0.0117	0.0150	
0.05	0.0550	0.0517	0.0617	0.0517	0.0500	0.0508	0.0467	0.25
0.10	0.1008	0.1025	0.1050	0.1000	0.0925	0.1050	0.1092	
P\\$\hat{\theta}$	0.3017	0.1500	0.1988	0.4026	0.3526	1,5481	1.0792	ρ_0
0.01	0.0117	0.0083	0.0108	0.0108	0.0133	0.0125	0.0133	
0.05	0.0533	0.0517	0.0633	0.0508	0.0500	0.0500	0.0458	0.3
0.10	0.0992	0.1033	0.1050	0.1017	0.0883	0.1025	0.1075	
P\\$\hat{\theta}$	0.3517	0.1500	0.1988	0.4025	0.3526	1.5481	1.0791	ρ_0
0.01	0.0117	0.0083	0.0108	0.0100	0.0142	0.0125	0.0125	
0.05	0.0517	0.0508	0.0625	0.0500	0.0525	0.0492	0.0458	0.35
0.10	0.0967	0.1033	0.1050	0.1017	0.0900	0.1000	0.1025	
P\\$\hat{\theta}$	0.4017	0.1051	0.1988	0.4024	0.3526	1.5482	1.0789	ρ_0
0.01	0.0125	0.0083	0.0108	0.0083	0.0142	0.0125	0.0100	
0.05	0.0542	0.0508	0.0625	0.0492	0.0517	0.0450	0.0500	0.4
0.10	0.0925	0.1042	0.1050	0.1042	0.0892	0.0950	0.1017	
P\\$\hat{\theta}$	0.4517	0.1501	0.1988	0.4024	0.3526	1.5482	1.0788	ρ_0
0.01	0.0125	0.0083	0.0100	0.0092	0.0133	0.0117	0.0075	
0.05	0.0525	0.0517	0.0633	0.0508	0.0508	0.0467	0.0508	0.45
0.10	0.0958	0.1042	0.1067	0.1025	0.0908	0.0958	0.0942	

Normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200} \sum_{j=1}^{1200} \theta_j^*$, $\sqrt{2.4} \simeq 1.5942$, $\sqrt{1.2} \simeq 1.0954$

Continue (b)

$$\begin{cases} \theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{2.4}, \sqrt{1.2}) \\ \theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{2.4}, \sqrt{1.2}) \end{cases}$$

P\\$\hat{\theta}$	0.5017	0.1501	0.1988	0.4023	0.3526	1.5483	1.0786	ρ_0
0.01	0.0117	0.0083	0.0100	0.0083	0.0142	0.0117	0.0033	
0.05	0.0500	0.0517	0.0642	0.0492	0.0517	0.0467	0.0492	0.5
0.10	0.0992	0.1042	0.1058	0.1075	0.0908	0.1025	0.0958	
P\\$\hat{\theta}$	0.5517	0.1501	0.1989	0.4022	0.3527	1.5484	1.0785	ρ_0
0.01	0.0133	0.0083	0.092	0.0083	0.0142	0.0100	0.0058	
0.05	0.0483	0.0517	0.0633	0.0500	0.0508	0.0483	0.0492	0.55
0.10	0.0942	0.1042	0.1067	0.1083	0.0908	0.0992	0.0983	
P\\$\hat{\theta}$	0.6016	0.1501	0.1989	0.4022	0.3527	1.5484	1.0784	ρ_0
0.01	0.0125	0.0083	0.0092	0.0067	0.0150	0.0100	0.0042	
0.05	0.0525	0.0517	0.0625	0.0500	0.0508	0.0483	0.0458	0.6
0.10	0.0933	0.1033	0.1067	0.1108	0.0917	0.0950	0.0983	
P\\$\hat{\theta}$	0.6515	0.1502	0.1989	0.4022	0.3528	1.5484	1.0787	ρ_0
0.01	0.0108	0.0092	0.0100	0.0067	0.0167	0.0100	0.0042	
0.05	0.0500	0.0517	0.0625	0.0500	0.0492	0.0475	0.0450	0.65
0.10	0.0958	0.1025	0.1067	0.1117	0.0908	0.0950	0.0950	
P\\$\hat{\theta}$	0.7016	0.1501	0.1990	0.4021	0.3529	1.5485	1.0779	ρ_0
0.01	0.0133	0.0092	0.0092	0.0042	0.0175	0.0125	0.0050	
0.05	0.0508	0.0508	0.0625	0.0533	0.0517	0.0492	0.0458	0.7
0.10	0.0975	0.1025	0.1050	0.1125	0.0900	0.0983	0.0925	

Normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{2.4} \simeq 1.5942$,$\sqrt{1.2} \simeq 1.0954$

Continue (c)

$$\begin{cases} \theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{2.4}, \sqrt{1.2}) \\ \theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{2.4}, \sqrt{1.2}) \end{cases}$$

P\\$\hat{\theta}$	0.7515	0.1501	0.1990	0.4022	0.3530	1.5485	1.0782	ρ_0
0.01	0.0142	0.0092	0.0100	0.0042	0.0175	0.0142	0.0058	
0.05	0.0558	0.0517	0.0600	0.0525	0.0492	0.0525	0.0417	0.75
0.10	0.0992	0.1025	0.1083	0.1100	0.0908	0.0950	0.0908	
P\\$\hat{\theta}$	0.8013	0.1501	0.1991	0.4024	0.3533	1.5484	1.0790	ρ_0
0.01	0.0125	0.0092	0.0100	0.0050	0.0617	0.0125	0.0050	
0.05	0.0542	0.0517	0.0592	0.0533	0.0483	0.0508	0.0433	0.8
0.10	0.1050	0.1017	0.1083	0.1058	0.0892	0.0967	0.0892	
P\\$\hat{\theta}$	0.8512	0.1501	0.1992	0.4026	0.3534	1.5483	1.0794	ρ_0
0.01	0.0117	0.0083	0.0100	0.0033	0.0167	0.0133	0.0067	
0.05	0.0542	0.0508	0.0592	0.0517	0.0492	0.0525	0.0450	0.85
0.10	0.1067	0.1008	0.1100	0.1075	0.0883	0.0992	0.0900	
P\\$\hat{\theta}$	0.9011	0.1501	0.1990	0.4022	0.3530	1.5485	1.0782	ρ_0
0.01	0.0100	0.0083	0.0092	0.0042	0.0167	0.0133	0.0075	
0.05	0.0583	0.0508	0.0583	0.0525	0.0492	0.0492	0.0458	0.9
0.10	0.1083	0.0992	0.1083	0.1058	0.0900	0.1033	0.0925	
P\\$\hat{\theta}$	0.9509	0.1501	0.1993	0.4030	0.3538	1.5480	1.0804	ρ_0
0.01	0.0092	0.0083	0.0108	0.0058	0.0167	0.0133	0.0092	
0.05	0.0608	0.0508	0.0592	0.0517	0.0475	0.0475	0.0458	0.95
0.10	0.1125	0.0983	0.1075	0.1083	0.0892	0.0892	0.0942	

Normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{2.4} \simeq 1.5942$,$\sqrt{1.2} \simeq 1.0954$

(d)

Table 2.7:

$$\begin{cases} \theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{2.4}, \sqrt{1.2}) \\ \theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{2.4}, \sqrt{1.2}) \end{cases}$$

P\ $\rho_0 \overset{=}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	ρ
0.01	0.0100	0.1083	0.5283	0.9000	0.9892	1.0000	1.0000	1.0000	
0.05	0.0583	0.2658	0.7617	0.9675	0.9975	1.0000	1.0000	1.0000	0
0.10	0.1058	0.3758	0.8358	0.9792	0.9983	1.0000	1.0000	1.0000	
P\ $\rho_0 \overset{=}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	ρ
0.01	0.5233	0.0983	0.0108	0.0958	0.4775	0.8833	0.9825	1.0000	
0.05	0.7483	0.2450	0.0567	0.2525	0.7217	0.9558	0.9958	1.0000	0.1
0.10	0.8392	0.3692	0.1033	0.3617	0.8117	0.9758	0.9975	1.0000	
P\ $\rho_0 \overset{=}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	ρ
0.01	0.9200	0.5033	0.0917	0.0100	0.0917	0.4675	0.8692	0.9800	
0.05	0.9708	0.7317	0.2342	0.0583	0.2508	0.7125	0.9517	0.9942	0.15
0.10	0.9908	0.8267	0.3567	0.1025	0.3525	0.7992	0.9733	0.9975	
P\ $\rho_0 \overset{=}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	ρ
0.01	0.9975	0.9075	0.4817	0.0867	0.0117	0.0892	0.4525	0.8508	
0.05	0.9992	0.9675	0.7133	0.2325	0.0567	0.2550	0.6908	0.9450	0.2
0.10	0.9992	0.9883	0.8092	0.3442	0.1017	0.3508	0.7808	0.9708	

Normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{2.4} \simeq 1.5942$, $\sqrt{1.2} \simeq 1.0954$

Continue (a)

$$\begin{cases} \theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{2.4}, \sqrt{1.2}) \\ \theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{2.4}, \sqrt{1.2}) \end{cases}$$

P\ $\rho_0 =$ →	0.1	0.15	0.2	0.25	0.3	0.35	0.45	0.5	ρ
0.01	0.8950	0.4633	0.0792	0.0125	0.0867	0.4342	0.8308	0.9750	
0.05	0.9642	0.6992	0.2250	0.0550	0.2542	0.6708	0.9367	0.9875	0.25
0.10	0.9833	0.7958	0.3367	0.1008	0.3425	0.7608	0.9650	0.9958	
P\ $\rho_0 =$ →	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	ρ
0.01	0.8725	0.4392	0.0767	0.0117	0.0875	0.4233	0.8117	0.9725	
0.05	0.9617	0.6825	0.2200	0.0533	0.2492	0.6483	0.9258	0.9858	0.3
0.10	0.9817	0.7925	0.3333	0.0992	0.3350	0.7492	0.9608	0.9942	
P\ $\rho_0 =$ →	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	ρ
0.01	0.8617	0.4167	0.0725	0.0117	0.0875	0.4175	0.7983	0.9675	
0.05	0.9583	0.6683	0.2150	0.0517	0.2417	0.6350	0.9167	0.9842	0.35
0.10	0.9792	0.7800	0.3200	0.0967	0.3308	0.7375	0.9583	0.9892	
P\ $\rho_0 =$ →	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	ρ
0.01	0.8450	0.3975	0.0617	0.0125	0.0883	0.4142	0.7775	0.9633	
0.05	0.9558	0.6550	0.2017	0.0542	0.2283	0.6267	0.0.9075	0.9825	0.4
0.10	0.9775	0.7700	0.3150	0.0925	0.3317	0.7275	0.9517	0.9867	
P\ $\rho_0 =$ →	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	ρ
0.01	0.8350	0.3825	0.0525	0.0125	0.0883	0.4050	0.7683	0.9550	
0.05	0.9517	0.6533	0.1975	0.0525	0.2292	0.6167	0.8942	0.9817	0.45
0.10	0.9750	0.7633	0.3150	0.0958	0.3333	0.7092	0.9400	0.9850	

Normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200} \theta_j^*$, $\sqrt{2.4} \simeq 1.5942$, $\sqrt{1.2} \simeq 1.0954$

Continue (b)

$$\begin{cases} \theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{2.4}, \sqrt{1.2}) \\ \theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{2.4}, \sqrt{1.2}) \end{cases}$$

P\ $\rho_0 \overset{=}{\rightarrow}$	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	ρ
0.01	0.8333	0.3725	0.0475	0.0117	0.0925	0.4067	0.7558	0.9483	
0.05	0.9550	0.6342	0.1867	0.0500	0.2258	0.6117	0.8833	0.9808	0.5
0.10	0.9758	0.7550	0.3067	0.0992	0.3283	0.7042	0.9308	0.9833	
P\ $\rho_0 \overset{=}{\rightarrow}$	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	ρ
0.01	0.8308	0.3658	0.0408	0.0133	0.0992	0.4017	0.7475	0.9383	
0.05	0.9517	0.6317	0.1833	0.0483	0.2375	0.6092	0.8808	0.9792	0.55
0.10	0.9767	0.7508	0.2925	0.0942	0.3350	0.7058	0.9233	0.9825	
P\ $\rho_0 \overset{=}{\rightarrow}$	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	ρ
0.01	0.8367	0.3658	0.0375	0.0125	0.1067	0.3967	0.7492	0.9300	
0.05	0.9525	0.6367	0.1833	0.0525	0.2450	0.6033	0.8800	0.9758	0.6
0.10	0.9783	0.7525	0.2833	0.0933	0.3358	0.7108	0.9267	0.9817	
P\ $\rho_0 \overset{=}{\rightarrow}$	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	ρ
0.01	0.8533	0.3708	0.0333	0.0108	0.1058	0.4008	0.7525	0.9308	
0.05	0.9592	0.6458	0.1817	0.0500	0.2517	0.6133	0.8850	0.9725	0.65
0.10	0.9792	0.7600	0.2900	0.0958	0.3417	0.7133	0.9275	0.9833	
P \ $\rho_0 \overset{=}{\rightarrow}$	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	ρ
0.01	0.8867	0.3867	0.0375	0.0133	0.1150	0.4233	0.7683	0.9325	
0.05	0.9675	0.6625	0.1792	0.0508	0.2625	0.6325	0.8892	0.9725	0.7
0.10	0.9817	0.7742	0.2975	0.0975	0.3492	0.7250	0.9292	0.9825	

Normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{2.4} \simeq 1.5942$, $\sqrt{1.2} \simeq 1.0954$

Continue (c)

$$\begin{cases} \theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{2.4}, \sqrt{1.2}) \\ \theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{2.4}, \sqrt{1.2}) \end{cases}$$

P\ $\rho_0 \overset{=}{\to}$	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	ρ
0.01	0.9033	0.4208	0.0383	0.0142	0.1258	0.4517	0.9375	0.9808	
0.05	0.9725	0.6842	0.1825	0.0558	0.2767	0.6458	0.9767	0.9942	0.75
0.10	0.9875	0.8017	0.3067	0.0992	0.3617	0.7425	1.0000	0.9958	
P\ $\rho_0 \overset{=}{\to}$	0.65	0.7	0.75	0.80	0.85	0.9	0.95	1	ρ
0.01	0.9333	0.4708	0.0425	0.0125	0.1392	0.4917	0.8267	0.9500	
0.05	0.9825	0.7350	0.2008	0.0542	0.2867	0.6875	0.9050	0.9783	0.8
0.10	0.9925	0.8392	0.3233	0.1050	0.3742	0.7733	0.9408	0.9858	
P \ $\rho_0 \overset{=}{\to}$	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	ρ
0.01	1.0000	0.9575	0.5367	0.0567	0.0117	0.1458	0.5492	0.8558	
0.05	1.0000	0.9917	0.7967	0.2217	0.0542	0.2967	0.7308	0.9250	0.85
0.10	1.0000	0.9967	0.8808	0.3458	0.1067	0.3983	0.7967	0.9550	
P \ $\rho_0 \overset{=}{\to}$	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	ρ
0.01	0.9992	0.9992	0.9758	0.6300	0.0742	0.0100	0.1617	0.6117	
0.05	1.0000	1.0000	0.9967	0.8575	0.2583	0.0583	0.3258	0.7750	0.9
0.10	1.0000	1.0000	1.0000	0.9142	0.3858	0.1083	0.4400	0.8308	
P \ $\rho_0 \overset{=}{\to}$	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	ρ
0.01	1.0000	1.0000	1.0000	0.9908	0.7217	0.1033	0.0092	0.1850	
0.05	1.0000	1.0000	1.0000	1.0000	0.9025	0.2942	0.0608	0.3717	0.95
0.10	1.0000	1.0000	1.0000	1.0000	0.9500	0.4333	0.1125	0.4933	

Normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200} \sum_{j=1}^{1200} \theta_j^*$, $\sqrt{2.4} \simeq 1.5942$, $\sqrt{1.2} \simeq 1.0954$

(d)

Table 2.8: $H_0 : \theta = \theta_0,\ where\ \rho = 0 \sim 0.95$

$$\begin{cases} \theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{1.5}, \sqrt{1.25}) \\ \theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{1.5}, \sqrt{1.25}) \end{cases}$$

P\\$\hat{\theta}$	7×10^{-4}	0.1498	0.2017	0.3986	0.3490	1.2220	1.1070	ρ_0
0.01	0.0108	0.0125	0.0108	0.0125	0.0133	0.0733	0.0200	
0.05	0.0433	0.0517	0.0558	0.0633	0.0558	0.1442	0.0850	0
0.10	0.0842	0.1058	0.0992	0.1125	0.0983	0.2283	0.0842	
P\\$\hat{\theta}$	0.0998	0.1504	0.2045	0.3977	0.3450	1.2219	1.1064	ρ_0
0.01	0.0100	0.0125	0.0100	0.0108	0.0117	0.0725	0.0200	
0.05	0.0450	0.0517	0.0525	0.0625	0.0533	0.1467	0.0833	0.1
0.10	0.0842	0.1058	0.1000	0.1100	0.0992	0.2250	0.1425	
P\\$\hat{\theta}$	0.1498	0.1504	0.2045	0.3977	0.3450	1.2219	1.1064	ρ_0
0.01	0.0083	0.0125	0.0100	0.0108	0.0125	0.0725	0.0192	
0.05	0.0483	0.0517	0.0517	0.0600	0.0533	0.1467	0.0808	0.15
0.10	0.0867	0.1058	0.0992	0.1117	0.0983	0.2242	0.1433	
P\\$\hat{\theta}$	0.1998	0.1504	0.2045	0.3977	0.3450	1.2219	1.1065	ρ_0
0.01	0.0100	0.0125	0.0092	0.0108	0.0125	0.0675	0.0183	
0.05	0.0467	0.0517	0.0525	0.0600	0.0558	0.1475	0.0800	0.2
0.10	0.0908	0.1058	0.0992	0.1100	0.0992	0.2242	0.1450	

Non-normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{1.5} \simeq 1.2247$,$\sqrt{1.25} \simeq 1.11870$

Continue (a)

$$\begin{cases} \theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{1.5}, \sqrt{1.25}) \\ \theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{1.5}, \sqrt{1.25}) \end{cases}$$

P \ $\hat{\theta}$	0.2499	0.1500	0.2042	0.3983	0.3451	1.2220	1.1074	ρ_0
0.01	0.0100	0.0125	0.0083	0.0108	0.0108	0.0575	0.0183	
0.05	0.0417	0.0567	0.0492	0.0658	0.0633	0.1367	0.0792	0.25
0.10	0.0825	0.1083	0.0925	0.1042	0.1058	0.2092	0.1350	
P\ $\hat{\theta}$	0.2998	0.1504	0.2045	0.3977	0.3451	1.2219	1.1180	ρ_0
0.01	0.0117	0.0125	0.0117	0.0108	0.0117	0.0608	0.0183	
0.05	0.0517	0.0517	0.0525	0.0608	0.0558	0.1475	0.0758	0.3
0.10	0.0950	0.1050	0.0983	0.1050	0.0.0975	0.2208	0.1383	
P\ $\hat{\theta}$	0.3497	0.1504	0.2045	0.3977	0.3451	1.2219	1.1066	ρ_0
0.01	0.0125	0.0125	0.0092	0.0125	0.0108	0.0575	0.0175	
0.05	0.0508	0.0517	0.0533	0.0575	0.0608	0.1442	0.0717	0.35
0.10	0.0925	0.1058	0.0975	0.1092	0.0958	0.2208	0.1325	
P\ $\hat{\theta}$	0.3997	0.1504	0.2046	0.3978	0.3451	1.2219	1.1068	ρ_0
0.01	0.0108	0.0125	0.0100	0.0125	0.0100	0.0575	0.0158	
0.05	0.0508	0.0517	0.0525	0.0575	0.0592	0.1433	0.0725	0.4
0.10	0.0983	0.1067	0.0975	0.1083	0.0950	0.2200	0.1325	
P\ $\hat{\theta}$	0.4497	0.1504	0.2046	0.3978	0.3452	1.2219	1.1069	ρ_0
0.01	0.0100	0.0125	0.0100	0.0133	0.0100	0.0550	0.0158	
0.05	0.0508	0.0508	0.0508	0.0550	0.0600	0.1400	0.0692	0.45
0.10	0.1017	0.1058	0.0967	0.1067	0.0958	0.2192	0.1275	

Non-normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{1.5} \simeq 1.2247$, $\sqrt{1.25} \simeq 1.11870$

Continue (b)

$$\begin{cases} \theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{1.5}, \sqrt{1.25}) \\ \theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{1.5}, \sqrt{1.25}) \end{cases}$$

P\\$\hat{\theta}$	0.4994	0.1507	0.2043	0.3984	0.3456	1.2219	1.1160	ρ_0
0.01	0.0050	0.0133	0.0092	0.0133	0.0108	0.0525	0.0158	
0.05	0.0533	0.0550	0.0517	0.0500	0.0533	0.1292	0.0575	0.5
0.10	0.0942	0.1000	0.0950	0.1000	0.0992	0.2092	0.1142	
P\\$\hat{\theta}$	0.5496	0.1504	0.2047	0.3979	0.3453	1.2218	1.1072	ρ_0
0.01	0.0100	0.0125	0.0117	0.0133	0.0108	0.0558	0.0158	
0.05	0.0550	0.0508	0.0508	0.0525	0.0542	0.1375	0.0667	0.55
0.10	0.1058	0.1042	0.1000	0.1100	0.0942	0.2158	0.1283	
P\\$\hat{\theta}$	0.5995	0.1504	0.2047	0.3980	0.3453	1.2218	1.1074	ρ_0
0.01	0.1050	0.1042	0.1000	0.1100	0.0950	0.2125	0.1242	
0.05	0.0550	0.0508	0.0517	0.0517	0.0550	0.1400	0.0633	0.6
0.10	0.1050	0.1042	0.1000	0.1100	0.0950	0.2125	0.1242	
P\\$\hat{\theta}$	0.6495	0.1504	0.2048	0.3982	0.3454	1.2218	1.1077	ρ_0
0.01	0.0125	0.0125	0.0083	0.0125	0.0108	0.0542	0.0158	
0.05	0.0567	0.0500	0.0508	0.0550	0.0525	0.1417	0.0658	0.65
0.10	0.1050	0.1033	0.0992	0.1108	0.0942	0.2125	0.1258	
P\\$\hat{\theta}$	0.6995	0.1504	0.2048	0.3983	0.3455	1.2217	1.1079	ρ_0
0.01	0.0142	0.0125	0.0083	0.0133	0.0117	0.0542	0.0142	
0.05	0.0567	0.0500	0.0508	0.0542	0.0525	0.1458	0.0650	0.7
0.10	0.1033	0.1050	0.0983	0.1092	0.0967	0.2133	0.1242	

Non-normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{1.5} \simeq 1.2247$, $\sqrt{1.25} \simeq 1.11870$

Continue (c)

$$\begin{cases} \theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{1.5}, \sqrt{1.25}) \\ \theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{1.5}, \sqrt{1.25}) \end{cases}$$

P\\$\hat{\theta}$	0.7494	0.1504	0.2049	0.3984	0.3455	1.2217	1.1081	ρ_0
0.01	0.0133	0.0125	0.0083	0.0117	0.0117	0.0533	0.0150	
0.05	0.0550	0.0500	0.0525	0.0550	0.0525	0.1442	0.0633	0.75
0.10	0.1050	0.1050	0.0975	0.1067	0.0967	0.2167	0.1208	
P\\$\hat{\theta}$	0.7994	0.1504	0.2049	0.3986	0.3456	1.2216	1.1083	ρ_0
0.01	0.0142	0.0125	0.0083	0.0125	0.0117	0.0542	0.0117	
0.05	0.0575	0.0508	0.0533	0.0558	0.0542	0.1442	0.0658	0.8
0.10	0.1000	0.1042	0.0983	0.1092	0.0983	0.2200	0.1175	
P\\$\hat{\theta}$	0.8494	0.1504	0.2050	0.3987	0.3456	1.2216	1.1084	ρ_0
0.01	0.0158	0.0125	0.0083	0.0108	0.0117	0.0542	0.0142	
0.05	0.0550	0.0508	0.0525	0.0550	0.0517	0.1450	0.0642	0.85
0.10	0.1017	0.1042	0.1000	0.1083	0.0975	0.2200	0.1225	
P\\$\hat{\theta}$	0.8994	0.1504	0.2050	0.3988	0.3457	1.2215	1.1085	ρ_0
0.01	0.0200	0.0125	0.0083	0.0133	0.1117	0.0550	0.0150	
0.05	0.0583	0.0500	0.0517	0.0567	0.0508	0.1442	0.0675	0.9
0.10	0.0992	0.1042	0.1008	0.1117	0.0983	0.2217	0.1192	
P\\$\hat{\theta}$	0.9494	0.1504	0.2050	0.3989	0.3457	1.2215	1.1085	ρ_0
0.01	0.0200	0.0125	0.0083	0.0125	0.0117	0.0542	0.0158	
0.05	0.0542	0.0508	0.0525	0.0558	0.0517	0.1508	0.0700	0.95
0.10	0.0958	0.1042	0.1008	0.1133	0.1000	0.2208	0.1175	

Non-normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{1.5} \simeq 1.2247, \sqrt{1.25} \simeq 1.1180$

(d)

Table 2.9: $H_0 : \theta = \theta_0$, where $\rho = 0 \sim 0.95$

$$\begin{cases} \theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{1.5}, \sqrt{1.25}) \\ \theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{1.5}, \sqrt{1.25}) \end{cases}$$

P \ $\rho_0 =$ →	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	ρ
0.01	0.0108	0.1433	0.6225	0.9558	0.9992	1.0000	1.0000	1.0000	
0.05	0.0433	0.3242	0.8183	0.9900	1.0000	1.0000	1.0000	1.0000	0
0.10	0.0842	0.4342	0.8958	0.9958	1.0000	1.0000	1.0000	1.0000	
P \ $\rho_0 =$ →	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	ρ
0.01	0.5892	0.0992	0.0100	0.1342	0.5900	0.9417	0.9992	1.0000	
0.05	0.8075	0.2825	0.0450	0.3058	0.7933	0.9825	0.9992	1.0000	0.1
0.10	0.8958	0.3492	0.0842	0.4167	0.8742	0.9917	1.0000	1.0000	
P \ $\rho_0 =$ →	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	ρ
0.01	0.9667	0.5800	0.0950	0.0083	0.1300	0.5808	0.9383	0.9983	
0.05	0.9883	0.8025	0.2742	0.0483	0.3067	0.7858	0.9808	0.9992	0.15
0.10	0.9950	0.8842	0.3875	0.0867	0.4092	0.8675	0.9900	1.0000	
P \ $\rho_0 =$ →	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	ρ
0.01	0.9992	0.9592	0.5675	0.0833	0.0100	0.1333	0.5650	0.9292	
0.05	1.0000	0.9875	0.7925	0.2667	0.0467	0.3050	0.7783	0.9750	0.2
0.10	1.0000	0.9950	0.8775	0.3842	0.0908	0.4017	0.8617	0.9867	

Non-normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200} \theta_j^*$, $\sqrt{1.5} \simeq 1.2247$, $\sqrt{1.25} \simeq 1.11870$

Continue (a)

$$\begin{cases} \theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{1.5}, \sqrt{1.25}) \\ \theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{1.5}, \sqrt{1.25}) \end{cases}$$

P \ $\rho_0 = \rightarrow$	0.1	0.15	0.2	0.25	0.3	0.35	0.45	0.5	ρ
0.01	0.9617	0.5508	0.0900	0.0100	0.1342	0.5442	0.9175	0.9967	
0.05	0.9892	0.7808	0.2533	0.0417	0.2993	0.7742	0.9733	1.0000	0.25
0.10	0.9942	0.8658	0.3958	0.0825	0.3925	0.8542	1.0000	1.0000	
P \ $\rho_0 = \rightarrow$	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	ρ
0.01	0.9992	0.9500	0.5458	0.0817	0.0117	0.1308	0.5417	0.9167	
0.05	1.0000	0.9867	0.7725	0.2450	0.0517	0.2975	0.7642	0.9725	0.3
0.10	1.0000	0.9942	0.8625	0.3842	0.0950	0.3992	0.8508	0.9817	
P\ $\rho_0 = \rightarrow$	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	ρ
0.01	0.9475	0.5442	0.0758	0.0125	0.1300	0.5325	0.9117	0.9875	
0.05	0.9883	0.7658	0.2408	0.0508	0.2917	0.7608	0.9717	0.9992	0.35
0.10	0.9942	0.8600	0.3858	0.0925	0.3983	0.8508	0.9783	0.9992	
P\ $\rho_0 = \rightarrow$	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	ρ
0.01	0.9983	0.9483	0.5358	0.0775	0.0100	0.1292	0.5308	0.9050	
0.05	1.0000	0.9875	0.7542	0.2383	0.0508	0.3000	0.7592	0.9683	0.4
0.10	1.0000	0.9967	0.8617	0.3817	0.1017	0.4008	0.8500	0.9792	
P\ $\rho_0 = \rightarrow$	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	ρ
0.01	0.5375	0.0767	0.0108	0.1283	0.5283	0.9075	0.9883	0.9992	
0.05	0.7608	0.2358	0.0508	0.2933	0.7583	0.9692	0.9975	1.0000	0.45
0.10	0.8592	0.3833	0.0983	0.3975	0.8533	0.9783	0.9992	1.0000	

Non-normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200} \sum_{j=1}^{1200} \theta_j^*$, $\sqrt{1.5} \simeq 1.2247$,$\sqrt{1.25} \simeq 1.11870$

Continue (b)

$$\begin{cases} \theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{1.5}, \sqrt{1.25}) \\ \theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{1.5}, \sqrt{1.25}) \end{cases}$$

P \ $\rho_0 \underset{\rightarrow}{=}$	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	ρ
0.01	0.9525	0.5425	0.0758	0.0050	0.1342	0.5425	0.9067	0.9875	
0.05	0.9908	0.7542	0.2375	0.0533	0.3150	0.7642	0.9683	0.9908	0.5
0.10	0.9983	0.8633	0.3775	0.0942	0.4083	0.8583	0.9783	0.9983	
P \ $\rho_0 \underset{\rightarrow}{=}$	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	ρ
0.01	0.9583	0.5642	0.0775	0.0100	0.1425	0.5608	0.9183	0.9892	
0.05	0.9892	0.7825	0.2483	0.0550	0.3208	0.7658	0.9675	0.9967	0.55
0.10	0.9958	0.8775	0.3833	0.1058	0.4050	0.8550	0.9817	0.9975	
P \ $\rho_0 \underset{\rightarrow}{=}$	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	ρ
0.01	0.9683	0.5883	0.0842	0.0125	0.1517	0.5917	0.9250	0.9900	
0.05	0.9917	0.8058	0.2583	0.0550	0.3300	0.7833	0.9692	0.9975	0.6
0.10	0.9942	0.8917	0.3842	0.1050	0.4192	0.8600	0.9825	0.9975	
P\ $\rho_0 \underset{\rightarrow}{=}$	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	ρ
0.01	0.9758	0.6308	0.1000	0.0125	0.1633	0.6250	0.9292	0.9925	
0.05	0.9933	0.8317	0.2800	0.0567	0.3392	0.7950	0.9750	0.9975	0.65
0.10	0.9950	0.9075	0.3900	0.1050	0.4392	0.8742	0.9875	0.9975	
P\ $\rho_0 \underset{\rightarrow}{=}$	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	ρ
0.01	0.9833	0.6725	0.1183	0.0142	0.1817	0.6658	0.9425	0.9967	
0.05	0.9950	0.8667	0.3008	0.0567	0.3525	0.8267	0.9808	0.9975	0.7
0.10	0.9967	0.9250	0.4150	0.1033	0.4692	0.8900	0.9917	0.9975	

Non-normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}\theta_j^*$, $\sqrt{1.5} \simeq 1.2247$,$\sqrt{1.25} \simeq 1.11870$

Continue (c)

$$\begin{cases} \theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{1.5}, \sqrt{1.25}) \\ \theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{1.5}, \sqrt{1.25}) \end{cases}$$

$P \setminus \overset{\rho_0}{\rightarrow}$	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	ρ
0.01	0.9900	0.7333	0.1367	0.0133	0.2042	0.7150	0.9975	1.0000	
0.05	0.9958	0.9025	0.3325	0.0550	0.3817	0.8550	0.9958	1.0000	0.75
0.10	0.9983	0.9450	0.4483	0.1017	0.4975	0.9025	0.9983	1.0000	

$P \setminus \overset{\rho_0 =}{\rightarrow}$	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	ρ
0.01	0.9942	0.8033	0.1617	0.0142	0.2258	0.7567	0.9800	0.9975	
0.05	0.9975	0.9300	0.3758	0.0575	0.4292	0.8817	0.9933	0.9983	0.8
0.10	0.9983	0.9633	0.5025	0.1000	0.5350	0.9400	0.9967	0.9992	

$P \setminus \overset{\rho_0 =}{\rightarrow}$	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	ρ
0.01	1.0000	0.9975	0.8700	0.1883	0.0158	0.2725	0.8192	0.9875	
0.05	1.0000	0.9983	0.9583	0.4175	0.0550	0.4675	0.9275	0.9967	0.85
0.10	1.0000	1.0000	0.9783	0.5500	0.1017	0.5808	0.9617	0.9983	

$P \setminus \overset{\rho_0 =}{\rightarrow}$	0.65	0.70	0.75	0.80	0.85	0.9	0.95	1	ρ
0.01	1.0000	1.0000	0.9983	0.9925	0.2483	0.0200	0.3183	0.8725	
0.05	1.0000	1.0000	1.0000	0.9792	0.4800	0.0583	0.5242	0.9575	0.9
0.10	1.0000	1.0000	1.0000	0.9900	0.6133	0.0992	0.6458	0.9775	

$P \setminus \overset{\rho_0 =}{\rightarrow}$	0.65	0.70	0.75	0.8	0.85	0.9	0.95	1	ρ
0.01	1.0000	1.0000	0.9625	0.3125	0.0200	0.3817	0.9342	0.9983	
0.05	1.0000	1.0000	0.9925	0.5667	0.0542	0.5942	0.9800	1.0000	0.95
0.10	1.0000	1.0000	0.9958	0.6967	0.0958	0.7033	0.9867	1.0000	

Non-normality

Repetitions=1200, $\hat{\theta} = \frac{1}{1200}\sum_{j=1}^{1200}, \theta_j$, $\sqrt{1.5} \simeq 1.2247$, $\sqrt{1.25} \simeq 1.1180$

(d)

Table 2.10: CMLE for the dynamic panel data of log-wage with unobserved heterogeneity period:1980 $\sim$ 1987

coefficient	$\hat{\rho}$	$\hat{\beta}$	$\hat{\alpha_0}$	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\sigma}_\varepsilon$	$\hat{\sigma}_a$
CMLE	0.3380	0.0474	0.8721	0.1745	0.0488	0.3506	0.2148
t-statistics	(18.330)	(2.174)	(25.251)	(8.224)	(1.253)	(77.473)	(18.897)

Table 2.11: Performance of DIF, GMM, GLS and CMLE (a)

(T=4, N=100)

ρ	Mean	GMM2 (DIF) RMSE SD	Mean	GMM2 (SYS) RMSE SD	Mean	GMM2 (ALL) RMSE SD
0.0	−0.0044	0.1227	0.0100	0.0994	0.0060	0.0970
		0.1227		0.0990		0.0969
0.3	0.2865	0.1853	0.3132	0.1221	0.3100	0.1216
		0.1849		0.1215		0.1213
0.5	0.4641	0.2693	0.5100	0.1333	0.5100	0.1356
		0.2674		0.1330		0.1353
0.8	0.4844	0.8805	0.8101	0.1620	0.8169	0.1541
		0.8824		0.1618		0.1533
0.9	0.2264	1.0659	0.9405	0.1615	0.9422	0.1415
		0.8264		0.1564		0.1351

ρ	Mean	CGLS RMSE SD	Mean	CMLE RMSE SD
0.0	0.0157	0.0986	0.0054	0.0916
		0.0974		0.0957
0.3	0.3188	0.1228	0.3054	0.1034
		0.1215		0.1067
0.5	0.5182	0.1353	0.5068	0.1082
		0.1342		0.1036
0.8	0.8365	0.1396	0.8004	0.0696
		0.1349		0.0684
0.9	0.9572	0.1121	0.8988	0.0355
		0.0964		0.0351

(a)

(T=4, N=200)

ρ	Mean	GMM2 (DIF) RMSE SD	Mean	GMM2 (SYS) RMSE SD	Mean	GMM2 (ALL) RMSE SD
0.0	−0.0037	0.0854	0.0051	0.0670	0.0028	0.0651
		0.0854		0.0669		0.0651
0.3	0.2919	0.1272	0.3092	0.0838	0.3061	0.0812
		0.1270		0.0833		0.0810
0.5	0.4828	0.1828	0.5098	0.0941	0.5079	0.0925
		0.1821		0.0936		0.0922
0.8	0.6362	0.5468	0.8050	0.1196	0.8112	0.1143
		0.5219		0.1195		0.1138
0.9	0.3731	1.1000	0.9235	0.1499	0.9308	0.1243
		0.9661		0.1481		0.1205

ρ	Mean	CGLS RMSE SD	Mean	CMLE RMSE SD
0.0	0.0083	0.0700	0.0007	0.0638
		0.0696		0.0654
0.3	0.3120	0.0895	0.3001	0.0708
		0.0887		0.0718
0.5	0.5135	0.1015	0.5068	0.1083
		0.1006		0.1036
0.8	0.8259	0.1115	0.8004	0.0696
		0.1085		0.0684
0.9	0.9431	0.1022	0.8999	0.0251
		0.0927		0.0249

(b)

(T=4, N=500)

ρ	Mean	GMM2 (DIF) RMSE SD	Mean	GMM2 (SYS) RMSE SD	Mean	GMM2 (ALL) RMSE SD
0.0	−0.0033	0.0557	0.0012	0.0434	0.0001	0.0421
		0.0556		0.0434		0.0442
0.3	0.2936	0.0827	0.3025	0.0552	0.3008	0.0530
		0.0824		0.0552		0.0530
0.5	0.4887	0.1177	0.5021	0.0632	0.5006	0.0612
		0.1172		0.0632		0.0612
0.8	0.7386	0.3144	0.7939	0.0781	0.7942	0.0770
		0.3085		0.0779		0.0769
0.9	0.5978	0.7081	0.9043	0.1000	0.9038	0.0884
		0.6401		0.0099		0.0883

ρ	Mean	CGLS RMSE SD	Mean	CMLE RMSE SD
0.0	0.0025	0.0462	0.0013	0.0406
		0.0461		0.0406
0.3	0.3030	0.0607	0.3022	0.0592
		0.0606		0.0588
0.5	0.5025	0.0710	0.5008	0.0441
		0.0710		0.0438
0.8	0.8007	0.0853	0.7999	0.0303
		0.0853		0.0309
0.9	0.9172	0.0880	0.8997	0.0158
		0.0863		0.0162

(c)

Table 2.12: Performance of DIF, GMM, GLS and CMLE (b)

(T=11, N=100)

ρ	$Mean$	$GMM2$ (DIF) $RMSE$ SD	$Mean$	$GMM2$ (SYS) $RMSE$ SD	$Mean$	$GMM2$ (ALL) $RMSE$ SD
0.0	−0.0138	0.0483	−0.0183	0.0468	−0.0153	0.0467
		0.0463		0.0431		0.0441
0.3	0.2762	0.0591	0.2728	0.0558	0.2795	0.0545
		0.0541		0.0487		0.0506
0.5	0.4629	0.0725	0.4689	0.0618	0.4794	0.0592
		0.0623		0.0535		0.0555
0.8	0.6812	0.1576	0.7925	0.0655	0.8043	0.0624
		0.1036		0.0651		0.0623
0.9	0.6455	0.2996	0.9259	0.0522	0.9302	0.0523
		0.1581		0.0453		0.0428

ρ	$Mean$	$CGLS$ $RMSE$ SD	$Mean$	$CMLE$ $RMSE$ SD
0.0	−0.0071	0.0364	0.0011	0.0350
		0.0358		0.0371
0.3	0.2832	0.0424	0.3005	0.0350
		0.0389		0.0365
0.5	0.4761	0.0490	0.5001	0.0328
		0.0428		0.0336
0.8	0.8025	0.0595	0.8002	0.0180
		0.0595		0.0179
0.9	0.9422	0.0623	0.8999	0.0076
		0.0459		0.0074

(a)

(T=11, N=200)

ρ	GMM2 (DIF) Mean	GMM2 (DIF) RMSE / SD	GMM2 (SYS) Mean	GMM2 (SYS) RMSE / SD	GMM2 (ALL) Mean	GMM2 (ALL) RMSE / SD
0.0	−0.0070	0.0358	−0.0059	0.0310	−0.0057	0.0313
		0.0352		0.0304		0.0307
0.3	0.2883	0.0427	0.2914	0.0345	0.2925	0.0348
		0.0411		0.0335		0.0340
0.5	0.4815	0.0503	0.4899	0.0373	0.4922	0.0373
		0.0468		0.0359		0.0365
0.8	0.7373	0.0971	0.8025	0.0421	0.8075	0.0430
		0.0742		0.0420		0.0423
0.9	0.7256	0.2152	0.9231	0.0435	0.9263	0.0445
		0.1261		0.0369		0.0359

ρ	CGLS Mean	CGLS RMSE / SD	CMLE Mean	CMLE RMSE / SD
0.0	−0.0037	0.0272	0.0006	0.0248
		0.0270		0.0262
0.3	0.2907	0.0318	0.3002	0.0248
		0.0304		0.0258
0.5	0.4858	0.0369	0.5000	0.0232
		0.0340		0.0238
0.8	0.8039	0.0449	0.8003	0.0127
		0.0448		0.0127
0.9	0.9345	0.0506	0.8999	0.0053
		0.0370		0.0053

(b)

(T=11, N=500)

ρ	Mean	GMM2 (DIF) RMSE SD	Mean	GMM2 (SYS) RMSE SD	Mean	GMM2 (ALL) RMSE SD
0.0	−0.0025	0.0201	−0.0010	0.0172	−0.0012	0.0173
		0.0200		0.0172		0.0173
0.3	0.2959	0.0237	0.2986	0.0182	0.2984	0.0183
		0.0233		0.0181		0.0182
0.5	0.4934	0.0276	0.4984	0.0189	0.4983	0.0190
		0.0268		0.0189		0.0190
0.8	0.7695	0.0536	0.8019	0.0244	0.8027	0.0249
		0.0441		0.0243		0.0248
0.9	0.8110	0.1168	0.9120	0.0306	0.9135	0.0312
		0.0757		0.0280		0.0282

ρ	Mean	CGLS RMSE SD	Mean	CMLE RMSE SD
0.0	−0.0016	0.0169	0.0003	0.0406
		0.0168		0.0410
0.3	0.2960	0.0196	0.3003	0.0442
		0.0192		0.0442
0.5	0.4937	0.0228	0.5000	0.0147
		0.0220		0.0143
0.8	0.8008	0.0316	0.8003	0.0080
		0.0316		0.0079
0.9	0.9206	0.0361	0.9002	0.0034
		0.0296		0.0033

(c)

Table 2.13: Performance of DIF, GMM, GLS and CMLE (c)

Estimator	Mean	Std. Dev.	Mean ASE	RMSE
OLS	0.8740	0.0203		0.3746
Within	-0.0343	0.0565		0.5373
GLS	0.6659	0.0965		0.1919
GMM(DIF)	0.4867	0.1844	0.1775	0.1848
GMM(SYS)	0.4999	0.1082	0.1068	0.1081
GMM(All)	0.5067	0.1109	0.1078	0.1111
CGLS	0.5124	0.1030		0.1037
CMLE	0.5179	0.1227		0.0769

N=200, T=4

Table 2.14: Performance of CMLE (d)

N,T	ρ	$\hat{\rho}$	Std. Dev.	MAE	RMSE
100,4	0.0	0.0119	0.1100	0.0837	0.0931
100,4	0.3	0.3217	0.1430	0.1052	0.1102
100,4	0.5	0.5322	0.1700	0.1228	0.1148
100,4	0.8	0.8116	0.1067	0.0821	0.0709
100,4	0.9	0.9027	0.0465	0.0366	0.0354
200,4	0.0	0.0065	0.0711	0.0837	0.0644
200,4	0.3	0.3155	0.0910	0.0697	0.0736
200,4	0.5	0.5179	0.1227	0.0816	0.0769
200,4	0.8	0.8078	0.0741	0.0573	0.0491
200,4	0.9	0.9021	0.0330	0.0259	0.0251
500,4	0.0	0.0027	0.0450	0.0359	0.0403
500,4	0.3	0.3043	0.0558	0.0445	0.0447
500,4	0.5	0.5057	0.0637	0.0501	0.0458
500,4	0.8	0.0823	0.0470	0.0983	0.0305
500,4	0.9	0.9003	0.0215	0.0173	0.0158

CHAPTER 3

Models Where State Dependence Depends On Unobserved Heterogeneity

3.1 Introduction

In the previous chapters we have discussed existing methods and the CMLE suggested by Wooldridge (2000b) for the AR(1) model with unobserved heterogeneity. The model is restrictive in that it assumes the amount of state dependence does not depend on unobserved heterogeneity. A more general model is

$$y_{it} = \rho \, y_{i,t-1} + a_i + \gamma \, (a_i \, y_{i,t-1}) + \varepsilon_{it} \qquad i = 1, \ldots, N, \; t = 1, \ldots, T, \tag{3.1}$$

which means that the amount of state dependence depends on the heterogeneity. The autoregressive coefficient for each i is given by $\rho + \gamma a_i$, so that it is a function of the individual heterogeneity. In (3.1) we clearly cannot estimate $\rho + \gamma a_i$ for each i with a short time period. But we can hopefully estimate the average effect, $\vartheta \equiv \rho + \gamma\mu_a$, where $\mu_a = E(a_i)$.

One interesting question surrounding model (3.1) is: Do standard IV methods applied to the first difference equation consistently estimate interesting parameters? To see that the answer is no, we consider the IV estimator that uses $\Delta y_{i,t-2}$ as a IV for $\Delta y_{i,t-1}$. Differencing (3.1) gives

$$\Delta y_{it} = \rho\, \Delta y_{i,t-1} + \gamma\, (a_i\, \Delta y_{i,t-1}) + \Delta\varepsilon_{it} \tag{3.2}$$

The IV estimator of ρ applied to the first differenced equation is, with fixed T, asymptotically equivalent to:

$$\begin{aligned}
&\operatorname*{plim}_{N\to\infty} \hat{\rho}_{IV} \\
&= \operatorname*{plim}_{N\to\infty} \frac{\frac{1}{NT}\sum_i\sum_t \Delta y_{it}\Delta y_{i,t-2}}{\frac{1}{NT}\sum_i\sum_t \Delta y_{i,t-1}\Delta y_{i,t-2}} \\
&= \operatorname*{plim}_{N\to\infty} \frac{\frac{1}{NT}\sum_i\sum_t(\rho\, \Delta y_{i,t-1} + \gamma\, (a_i\, \Delta y_{i,t-1}) + \Delta\varepsilon_{it})\Delta y_{i,t-2}}{\frac{1}{NT}\sum_i\sum_t \Delta y_{i,t-1}\Delta y_{i,t-2}} \\
&= \rho + \frac{\gamma\, \operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_i\sum_t a_i\, \Delta y_{i,t-1}\Delta y_{i,t-2}}{\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_i\sum_t \Delta y_{i,t-1}\Delta y_{i,t-2}} + \\
&\qquad \frac{\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_i\sum_t \Delta\varepsilon_{it}\Delta y_{i,t-2}}{\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_i\sum_t \Delta y_{i,t-1}\Delta y_{i,t-2}} \\
&= \rho + \gamma\mu_a + \frac{\gamma\, \operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_i\sum_t h_i\, \Delta y_{i,t-1}\Delta y_{i,t-2}}{\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_i\sum_t \Delta y_{i,t-1}\Delta y_{i,t-2}} + \\
&\qquad \frac{\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_i\sum_t \Delta\varepsilon_{it}\Delta y_{i,t-2}}{\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_i\sum_t \Delta y_{i,t-1}\Delta y_{i,t-2}},
\end{aligned} \tag{3.3}$$

where $h_i \equiv a_i - \mu_a$. To simplify the exposition, we make the following standard assumptions:

$$\begin{aligned}
E(\varepsilon_{it}|y_{i,t-1},\ldots,y_{i0},a_i) &= 0 && \text{for all } i,\ t{=}1,\ldots,T \\
\operatorname{Var}(\varepsilon_{it}|y_{i,t-1},\ldots,y_{i0},a_i) &= \sigma_\varepsilon^2 && \text{for all } i,\ t{=}1,\ldots,T
\end{aligned} \tag{3.4}$$

Obviously, $\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_i\sum_t \Delta\varepsilon_{it}\Delta y_{i,t-2}$ is zero under the above assumptions, but $\gamma \operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_i\sum_t h_i\,\Delta y_{i,t-1}\Delta y_{i,t-2}$ is equal to:

$$
\begin{aligned}
&\gamma \operatorname*{plim}_{N\to\infty} \tfrac{1}{NT}\textstyle\sum_i\sum_t h_i\,\Delta y_{i,t-1}\Delta y_{i,t-2} \\
&\quad = \gamma\left(\operatorname*{plim}_{N\to\infty} \tfrac{1}{N}\textstyle\sum_i h_i + \operatorname*{plim}_{N\to\infty} \tfrac{1}{N}\sum_i h_i \mathrm{Cov}_i(\Delta y_{i,t-1}, \Delta y_{i,t-2})\right) \\
&\quad = \gamma\tfrac{1}{T}\textstyle\sum_t E(h_i\Delta y_{i,t-1}, \Delta y_{i,t-2}),
\end{aligned}
\tag{3.5}
$$

and this does not equal zero without some unusual assumptions. Therefore, $\Delta y_{i,t-2}$ is not a valid IV for $\Delta y_{i,t-1}$ because $\Delta y_{i,t-2}$ and the error, $\gamma a_i\Delta y_{it-1} + \Delta\varepsilon_{it}$, are correlated and we are doing fixed T asymptotics.

We can also ask whether the IV estimator consistently estimates the average autoregressive coefficient, $\vartheta = \rho + \gamma\mu_a$. Unfortunately, the answer is no.

$$
\begin{aligned}
\operatorname*{plim}_{N\to\infty} \hat{\vartheta}_{IV} &= \frac{\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \Delta y_{i,t}\Delta y_{i,t-2}}{\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \Delta y_{i,t-1}\Delta y_{i,t-2}} \\
&= \frac{\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T ((\rho+\gamma a_i)\Delta y_{i,t-1} + \Delta\varepsilon_{it})\Delta y_{i,t-2}}{\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \Delta y_{i,t-1}\Delta y_{i,t-2}} \\
&= \frac{\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T ((\vartheta+\gamma h_i)\Delta y_{i,t-1} + \Delta\varepsilon_{it})\Delta y_{i,t-2}}{\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \Delta y_{i,t-1}\Delta y_{i,t-2}} \\
&= \vartheta + \frac{\gamma \operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T h_i\Delta y_{i,t-1}\Delta y_{i,t-2}}{\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \Delta y_{i,t-1}\Delta y_{i,t-2}} + \\
&\qquad \frac{\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \Delta\varepsilon_{it}\Delta y_{i,t-2}}{\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \Delta y_{i,t-1}\Delta y_{i,t-2}},
\end{aligned}
\tag{3.6}
$$

where $h_i = a_i - \mu_a$. The consistency of $\hat{\vartheta}_{IV}$ depends on the second term of the last expression of (3.6), equal to (3.5), which does not vanish even though we strengthen the assumption (3.4) such that h_i and ε_{it} are independent for all i, $t = 1, \ldots, T$.

The denominator of (3.6) is easily proved to be nonzero. The probability limit of the numerator of the third term on the right hand of (3.6) is zero, but the probability limit of the numerator of the second term is not always zero and thus the asymptotic bias of the IV estimator is given by the probability limit of the second term on the right hand of (3.6). Because the presence of the state dependence which depends on the unobserved effects, the unobserved effects are transmitted into the estimate of ρ. This causes that ρ can not be identified from $\rho + \gamma a_i$. In addition, $\hat{\vartheta}_{IV}$ cannot consistently estimate ϑ even if we further assume that c_i is independent of y_{i0}. Therefore, in the model with an interaction effect, the IV estimator does not consistently estimate ρ or the average autoregressive coefficient. It does not appear that differencing alone is a reasonable strategy for estimating the model with the interaction term.

Transformations other than differencing may work to estimate the parameters of (3.1), but they do not immediately suggest themselves. Instead, we can apply conditional MLE, as in the simpler model from Chapter 2. As in Chapter 2, we directly model the distribution of a_i given y_{i0} and any strictly exogenous variables. The most general model we consider is

$$\begin{aligned} y_{it} = \rho\, y_{i,t-1} + x_{it}\beta + a_i + \gamma\, [a_i\, y_{i,t-1}] + \varepsilon_{it} \\ i = 1, \ldots, N,\ t = 1, \ldots, T. \end{aligned} \tag{3.7}$$

We could consider a model where a_i also interacts with x_{it}, but the computational requirements would be severe. For many purposes, the most interest is in a model where unobserved heterogeneity interacts with the lagged dependent variable. The plan of this chapter is as follows. Section 2 considers model (3.1) and then considers model (3.7). Section 3 presents the simulation of CMLE for the model (3.1) and (3.7). Section 4 applies the models to log hourly wage for the panel of working man used in Chapter 2 in considering the interaction term, product of log hourly wage and the unobserved heterogeneity. Finally, Section 5 contains some concluding remarks.

3.2 AR(1) Models With Unobserved Heterogeneity, State Dependence

3.2.1 AR(1) Model Without Exogenous Variables

This section characterizes the CMLE for model (3.1). When i is omitted, we refer to a general cross-sectional observation. Using the general treatment in Chapter 2, the conditional densities corresponding to equation (3.1) is $f(y_t|y_{t-1}, a; \delta_0)$. We assume $D(\varepsilon_{it}|y_{i,t-1}, \ldots, y_{i0}, a_i) = D(\varepsilon_{it})$ and thus the joint density of $(y_T, \ldots, y_1)$ given $(y_0$, a) is

$$p(y_T, \ldots, y_1|y_0, a; \delta_0) = \prod_{t=1}^{T} f(y_t|y_{t-1}, a; \delta_0). \tag{3.8}$$

We can not estimate δ_0 by directly using (3.8) because it depends on a which is unobserved. According to the discussion of Chapter 2, we can model $D(a|y_0)$ and then construct the density $(y_T, \ldots, y_1)$ given y_0 by integrating out a from the joint density function. In practice, we can specify a parametric density :

$$h(a|y_0; \lambda_0), \tag{3.9}$$

where λ_0 is a vector of nuisance parameters. Wooldridge (2000b) suggested that we assume the a_i are from a conditional normal distribution, where the mean and variance given y_0 are flexible functions of y_0. Let us make assumptions on the ε_{it} and a_i as follows:

Assumption 3.1 $\varepsilon_{it}|y_{i,t-1}, \ldots, y_{i0}, a_i \sim \text{Normal}(0, \sigma_\varepsilon^2)$.

Assumption 3.2 $a_i|y_{i0} \sim \text{Normal}(\mu_a(y_{i0}), \sigma_a^2)$.

With a linear mean in Assumption 3.2 we can work $a_i = \alpha_0 + \alpha_1 y_{i0} + c_i$, where

c_i given y_{i0} is Normal(0 , σ_a^2). To characterize the conditional mean of y_{it} given y_{i0}, we use equation (3.1) to obtain

$$y_{it} = (\rho + \gamma\ a_i)^t y_{i0} + \left(\frac{1-(\rho+\gamma\ a_i)^t}{1-(\rho+\gamma\ a_i)}\right) a_i + \sum_{j=1}^{t}(\rho+\gamma\ a_i)^{j-1}\varepsilon_{i,t-j+1}. \tag{3.10}$$

Let us define a polynomial of t-order as $P_t(z) = w_{t,0} + w_{t,1}z + \ldots + w_{t,t-1}z^t$, where $P_0(z) \equiv w_{0,0}$, and $P_{-1} \equiv 0$. Then equation (3.10) can be rewritten as follows:

$$\begin{aligned} y_{it} \ = \ & [d^t + P_{t-1}(c_i)c_i]\cdot y_{i0} + \left(\textstyle\sum_{j=1}^{t}(d^{j-1} + P_{j-2}(c_i)c_i)\right)(\mu_a + c_i) \\ & + \textstyle\sum_{j=1}^{t}(d^{j-1} + P_{j-2}(c_i)c_i)\,\varepsilon_{i,t-j+1}, \end{aligned} \tag{3.11}$$

where $d = \rho + \gamma\ \mu_a$, and $w_{t,t-1} = \gamma^{t-1}$. From equation (3.11), the coefficient $w_{t,t-1}$ is a function of ρ, α_0, α_1, and y_{i0} over t from 1 to T: $w_{t,0} = d^t$ and $w_{t,t-1} = \gamma^t$. Under Assumption 3.1, ε_{it} is independent of y_{i0} for all t. We obtain $\mu(y_{i0}) \equiv E(y_{it}|y_{i0})$ as follows:

$$\begin{aligned} \mu(y_{i0}) = \ & d^t y_{i0} + \left(\frac{1-d^t}{1-d}\right)\mu_a + y_{i0}\textstyle\sum_{j=1}^{t} E(P_{j-1}(c_i)c_i|y_{i0}) + \\ & \mu_a \textstyle\sum_{j=1}^{t} E(P_{j-2}(c_i)c_i|y_{i0}) + \sum_{j=1}^{t} E(P_{j-2}(c_i)c_i^2|y_{i0}). \end{aligned} \tag{3.12}$$

Because we assume that $c_i|y_{i0}$ is Normal(0,σ_a^2), the t-th moment of c_i exists and is a function of σ_a. Therefore, the last three terms of equation (3.12), $\sum_{j=1}^{t} E(P_{j-1}(c_i)c_i|y_{i0})$, $\sum_{j=1}^{t} E(P_{j-2}(c_i)c_i|y_{i0})$, and $\sum_{j=1}^{t} E(P_{j-2}(c_i)c_i^2|y_{i0})$ can be concisely expressed as a function of ρ, α_0, α_1, y_{i0}, and σ_a. Therefore, $\mu(y_{i0}) = d^t y_{i0} + \left(\frac{1-d^t}{1-d}\right)\mu_a + \mathrm{A}(\alpha_0, \alpha_1, y_{i0}, \sigma_a)$. The conditional variance matrix $\Omega(y_{i0}) \equiv V(y_i|y_{i0})$ can be expressed as follows:

$$\begin{aligned} V(y_i|y_{i0}) = \ & E\left[((y_i - \mu(y_{i0})\,\ell_T)\,(y_i - \mu(y_{i0})\,\ell_T)'|y_{i0}\right] \\ = \ & E(y_i y_i'|y_{i0}) \ - \ \mu(y_{i0})^2 \ell_T\,\ell_T', \end{aligned} \tag{3.13}$$

Under Assumption 3.1 and 3.2, we can solve out the elements of the conditional variance matrix

$$\Omega(y_{i0}) = \begin{pmatrix} \omega_{11} & \ldots & \omega_{1T} \\ \vdots & \ddots & \vdots \\ \omega_{T1} & \ldots & \omega_{TT} \end{pmatrix}, \tag{3.14}$$

where ω_{st} is a function of ρ, α_0, α_1, y_{i0}, σ_a and σ_ε. In principle, we can solve out the formula for $\varepsilon(y_{i0})$ ($\equiv y_i - \mu(y_{i0})\,\ell_T$) and ω_{st} in terms of ρ, α_0, α_1, y_{i0}, σ_a and σ_ε based on (3.10) and (3.12). We set up the log-likelihood function away from the constant for cross-section observation i is as follows:

$$l(y_i;\theta) = -\frac{1}{2}\log(|\Omega(y_{i0})|) - \frac{1}{2}\left(\varepsilon(y_{i0})\Omega(y_{i0})^{-1}\varepsilon(y_{i0})'\right) \qquad (3.15)$$

Therefore, we can obtain the CMLE estimators by solving out the following maximizing problem:

$$\max_{\theta_0}\sum_{i=1}^{N} l(y_i;\theta) = \max_{\theta_0}\sum_{i=1}^{N}\left[-\frac{1}{2}\log(|\Omega(y_{i0})|) - \frac{1}{2}\left(\varepsilon(y_{i0})\Omega(y_{i0})^{-1}\varepsilon(y_{i0})'\right)\right] \qquad (3.16)$$

In fact, the complexity of the $\Omega(y_{i0})$ will increase with the value of T, so it become intractable for handling for equation (3.16). The general approach is that $f(Y_T|y_0, a)$ can be expressed as the product of $f(y_t|y_{t-1}, a)$ over t from 1 to T by parameterizing (3.8). And then we specify a conditional distribution of the unobserved effects $h(a|y_0)$, in particular the normal density function, $\frac{1}{\sqrt{2\pi\sigma_a^2}}\exp(\frac{-1}{2}(\frac{a-\mu_a}{\sigma_a})^2)$. We can set up the log density function as

$$\begin{aligned} l(Y_T;\theta) &= \log\int_{-\infty}^{\infty} f(Y_T|y_0, a)h(a|y_0)\mathrm{d}\,a \\ &= \log\int_{-\infty}^{\infty}\left[\prod_{t=1}^{T} f(y_t|y_{t-1}, a)\right] h(a|y_0)\mathrm{d}\,a \end{aligned} \qquad (3.17)$$

By equation (3.1) and Assumption 3.1, we obtain the following equation:

$$f(y_{it}|y_{i,t-1}, a) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\exp\left(\frac{-1}{2}\left[\frac{y_{it} - (\rho\, y_{i,t-1} + a_i + \gamma\,(a_i\, y_{i,t-1}))}{\sigma_\varepsilon}\right]^2\right) \qquad (3.18)$$

Putting (3.18) into (3.17) gives the log-likelihood function for each i is as

$$l(y_i;\theta) =$$

$$\log\int_{-\infty}^{\infty}\left(\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\right)^T\left[\prod_{t=1}^{T}\exp\left(\frac{-1}{2}\left[\frac{y_{it} - (\rho\, y_{i,t-1} + a + \gamma\,(a\, y_{i,t-1}))}{\sigma_\varepsilon}\right]^2\right)\right]\cdot$$

$$(\frac{1}{\sqrt{2\pi\sigma_a^2}})\exp(\frac{-1}{2}(\frac{a-\mu_a(y_{i0})}{\sigma_a})^2)\,\mathrm{d}\,a. \qquad (3.19)$$

We can obtain the CMLE estimators by maximizing the sum of function (3.19) across i from 1 to N.

3.2.2 AR(1) Model With Strictly Exogenous Variable

In this section we consider (3.8). Assuming strict exogeneity of x_t, we have

$$D(y_t|Y_{t-1}, X_T, a) = D(y_t|y_{t-1}, x_t, a), t = 1, \ldots, T. \tag{3.20}$$

Equation (3.20) means that once current x_t, and y_{t-1} and a are controlled for, x_s, $s \neq t$, do not affect the distribution of y_t. The conditional distribution can be parameterized as a conditional density,

$$f(y_t|y_{t-1}, x_t, a; \delta_0), \tag{3.21}$$

where the parameter δ_0 is finite dimensional parameters. In our application we assume that $f(y_t|Y_{t-1}, x_t, a; \delta_0)$ depends only on one lag of y_t and the current x_t. By the usual product of law for conditional densities, the joint density of $(y_T, \ldots, y_0)$ given $(x_T, \ldots, x_1, a)$ is as follows:

$$f(y_T, \ldots, y_1|x_T, \ldots, x_1, y_0,\ a\ ; \delta_0) = \prod_{t=1}^{T} f(y_t|y_{t-1}, x_t, a; \delta_0), \quad t = 1, \ldots, T. \tag{3.22}$$

As discussed in previous chapter, because the density of $(y_T, \ldots, y_1)$ given (x_T, ..., x_1, y_0, a) depends on a, which is unobserved, to consistently estimate δ_0, we integrate a out of the density. The recommended solution (Wooldridge [2000b]) is that to model a conditional distribution $D(a|X_T, y_0)$, and then construct the density of $(y_T, \ldots, y_1)$ given $(x_T, \ldots, x_1, y_0)$. It is crucial that this allows y_0 to be random and need not find, or even approximate, $D(y_0|X_T, a)$. Further, we do not have to specify an additional model for $D(a|X_T)$ or, assume that a and X_T are independent and then model D(a). In practice, we parameterize the conditional density : let

$$h(a|X_T, y_0; \lambda_0), \tag{3.23}$$

be the density corresponding to $D(a|X_T, y_0)$, where λ_0 is a vector of parameters. It is convenient to assume normality with conditional mean and variance in terms of (X_T, y_0). We make the following assumptions:

Assumption 3.3 $\varepsilon_{it}|Y_{i,t-1}, X_{i,T}, a_i \sim \text{Normal}(0, \sigma_\varepsilon^2)$.

Assumption 3.4 $a_i|x_i, y_{i0} \sim \text{Normal}(\mu_a(y_{i0}, \overline{x_i}), \sigma_a^2)$, where $\overline{x_i} = \frac{1}{T}\sum_{t=1}^{T} x_{it}$.

According to Assumption 3.4, we can write the equation for a_i as follows

$$a_i = \alpha_0 + \alpha_1 y_{i0} + \overline{x_i}\,\alpha_2 + c_i, i = 1, \ldots, N, \tag{3.24}$$

where c_i given (x_i, y_{i0}) is Normal(0 , σ_a^2). Assumption 3.4 and (3.24) imply that

$$h(a|X_T, y_0; \lambda_0) = \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left[\frac{-1}{2}\left(\frac{a_i - (\alpha_0 + \alpha_1 y_{i0} + \overline{x_i}\,\alpha_2)}{\sigma_a}\right)^2\right], \tag{3.25}$$

where $\lambda_0 = (\alpha_0, \alpha_1, \alpha_2, \sigma_a)$. Once we have specified $h(a|z_T, y_0, x_0; \lambda_0)$, we obtain the log-likelihood function for each cross section i as follow:

$$\begin{aligned} l(y_i, x_i; \theta) = \; & \log \int_{-\infty}^{\infty} \textstyle\prod_{t=1}^{T} f(y_{it}|x_{it}, y_{i,t-1}, a\; ; \delta_0)\cdot \\ & \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left[\frac{-1}{2}\left(\frac{a - (\alpha_0 + \alpha_1 y_{i0} + \overline{x_i}\,\alpha_2)}{\sigma_a}\right)^2\right] \mathrm{d}\,a, \end{aligned} \tag{3.26}$$

where θ_0 is a vector of all parameters of the model.

Under Assumption 3.3 and 3.4, we still can apply the procedures of the previous section to solve the CMLE. (3.11) and (3.12) can be re-written respectively as follows:

$$\begin{aligned} y_{it} = \; & [d^t + P_{t-1}(c_i)c_i]\, y_{i0} + \left(\textstyle\sum_{j=1}^{t}(d^{j-1} + P_{j-2}(c_i)c_i)\right)(\mu_a + c_i) \\ & + \textstyle\sum_{j=1}^{t}(d^{j-1} + P_{j-2}(c_i)c_i)\,\beta\, x_{i,t-j+1} + \sum_{j=1}^{t}(d^{j-1} + P_{j-2}(c_i)c_i)\,\varepsilon_{i,t-j+1}, \end{aligned} \tag{3.27}$$

and

$$E(y_{it}|X_{it},y_{i0}) = d^t y_{i0} + \left(\frac{1-d^t}{1-d}\right)\mu_a + y_{i0}\sum_{j=1}^{t} E(P_{j-1}(c_i)c_i|X_{it},y_{i0}) + \sum_{j=1}^{t}(d^{j-1} + E(P_{j-2}(c_i)c_i|X_{it},y_{i0}))\,\beta\, x_{i,t-j+1} + \mu_a\sum_{j=1}^{t} E(P_{j-2}(c_i)c_i|X_{it},y_{i0}) + \sum_{j=1}^{t} E(P_{j-2}(c_i)c_i^2|X_{it},y_{i0}). \tag{3.28}$$

Because we assume that $(c_i|x_i,y_{i0})$ is Normal(0,σ_a^2), the t-th moment of c_i exists and is a function of σ_a and hence the last three terms of equation (3.28): $\sum_{j=1}^{t} E(P_{j-1}(c_i)c_i|X_{it},y_{i0})$, $\sum_{j=1}^{t} E(P_{j-2}(c_i)c_i|X_{it},y_{i0})$, and $\sum_{j=1}^{t} E(P_{j-2}(c_i)c_i^2|X_{it},y_{i0})$ can be compressed as a function of ρ, α_0, α_1, $\overline{x_i}$, y_{i0}, and σ_a. Therefore, the $E(y_{it}|X_{it},y_{i0}) = d^t y_{i0} + \left(\frac{1-d^t}{1-d}\right)\mu_a + \sum_{j=1}^{t}(d^{j-1} + E(P_{j-2}(c_i)c_i|X_{it},y_{i0}))\,\beta\, x_{i,t-j+1} + \mathrm{A}(\alpha_0,\ \alpha_1,\ \overline{x_i},\ y_{i0},\ \sigma_a)$. If T is very small, say $T \leq 3$, we can use the log-likelihood function (3.15) and replace $\varepsilon(y_{i0})$ and $\Omega(y_{i0})$ with $\varepsilon(y_{i0},x_i)$ and Ω_{y_{i0},x_i} and then to maximize the sum of the log-likelihood function across i from 1 to N. If T is not very small, we use the following as log-likelihood function:

$$l(y_i,x_i;\theta) = \log\int_{-\infty}^{\infty}\left(\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\right)^T\left[\prod_{t=1}^{T}\exp\left(\frac{-1}{2}\left[\frac{y_{it}-(\rho\, y_{i,t-1}+x_{i,t}\beta+a+\gamma\,(a\,y_{i,t-1}))}{\sigma_\varepsilon}\right]^2\right)\right]\cdot \left(\frac{1}{\sqrt{2\pi\sigma_a^2}}\right)\exp\left(\frac{-1}{2}\left(\frac{a-\mu_a(y_{i0},\overline{x}_i)}{\sigma_{a_i}}\right)^2\right)\mathrm{d}\,a, \tag{3.29}$$

where $\mu_a(y_{i0},\overline{x}_i) = \alpha_0 + \alpha_1\, y_{i0} + \overline{x}_i\alpha_2$

By maximizing the sum of $l(y_i,x_i;\theta)$ over i from 1 to N, we can obtain the CMLE estimator. According to the discussion of the consistency in Section 2.2.2, the consistency of CMLE estimators that use (3.28) as the log-likelihood function can be ensured. If we have random sampling in the cross section dimension and standard regularity conditions, with fixed T, the CMLE for θ_0 will be consistent and

$\sqrt{N}$-asymptotically normally distributed. (See Newey and McFadden [1994] for sufficient regularity conditions.)

3.3 Simulation Evidence

3.3.1 Model without Exogenous Variable

The true model without exogenous variables of the simulation is as follows:

$$\begin{aligned} y_{it} &= \rho\, y_{i,t-1} + a_i + \gamma\, [a_i\, y_{i,t-1}] + \varepsilon_{it}, \\ i &= 1, \ldots, 250, \quad t = 1, \ldots, 5, \end{aligned} \tag{3.30}$$

where $a_i = \alpha_0 + \alpha_1\, y_{i0} + c_i$. The true value of $\theta = (\rho,\ \alpha_0,\ \alpha_1,\ \gamma,\ \sigma_\varepsilon,\ \sigma_a)$=($\rho$, 0.2, 0.4, γ, $\sqrt{2.4}$, $\sqrt{1.2}$), and we set ρ as different values equal to 0, 0.25, 0.5, 0.75, 0.9, 0.95 and γ as values equal to 0, 0.1. The values of c_i and ε_{it} are generated by N(0,1.2) and N(0,2.4), respectively. Because this method allows y_{i0} to be random, we generate the y_{i0} from the N(0,1). Using the generated data, we maximize the sum of log likelihood function (3.19) over i from 1 to N

By the use of the data coming from the above rule, we apply the MLE procedure of Gauss software to obtain the conditional maximum likelihood estimators. it is difficult and time-consuming to directly maximize the objective function , the sum of (3.19) over i from 1 to N, so we need find another easier numerical form for the log-likelihood function (3.19). We might calculate the integral of equation (3.19) by applying the formula for the evaluation of the necessary integral which is the Hermite integral formula $\int_{-\infty}^{\infty} \mathrm{e}^{-z^2} g(z) \mathrm{d}\, z = \sum_{j=1}^{K} w_j g(z_j)$, where K is the number of evaluation points, w_j is the weight given to the jth evaluation point, and $g(z_j)$ is $g(z)$ evaluated at the jth point of z (See Butler and Moffitt [1982]). Equation (3.19) can be re-written as follows:

$l(y_i;\theta) =$

$$\frac{-T-1}{2}\log 2\pi + \log\int_{-\infty}^{\infty}\left(\frac{1}{\sqrt{\sigma_\varepsilon^2}}\right)^T \left[\prod_{t=1}^T \exp\left(\frac{-1}{2\sigma_\varepsilon}(y_{it} - (\rho\ y_{i,t-1} + \sqrt{2}\sigma_a z + \mu_a(y_{i0}) + \gamma\ (\sqrt{2}\sigma_a z + \mu_a(y_{i0}))y_{i,t-1}))\right)\right]\exp(-z^2)\mathrm{d}\,z. \tag{3.31}$$

We can let $g(z)$ to be

$$\left(\frac{1}{\sqrt{\sigma_\varepsilon^2}}\right)^T \left[\prod_{t=1}^T \exp\left(\frac{-1}{2\sigma_\varepsilon}(y_{it} - (\rho\ y_{i,t-1}\sigma_a z + \mu_a(y_{i0}) + \gamma\ (\sqrt{2}\sigma_a z + \mu_a(y_{i0}))y_{i,t-1}))\right)\right]$$

and then the log-likelihood function away from the constant is

$$\log\int_{-\infty}^{\infty} g(z)\mathrm{e}^{-z^2}\mathrm{d}\,z \tag{3.32}$$

This formula is appropriate to our problem because the integration of equation (3.19) can be transformed as the product of $g(z)$ and e^{-z^2} equation. We can approximate the objective function $\sum_{i=1}^N \log\int_{-\infty}^{\infty} g(z)\mathrm{e}^{-z^2}\mathrm{d}\,z$ as follows

$$\sum_{i=1}^N \log\left(\sum_{j=1}^K w_j g(z_{ij})\right) \tag{3.33}$$

In the simulations we obtain the conditional maximum likelihood estimators from equation (3.33). The feasible computation of the Hermite integral depends on the number of evaluation points at which the integrand must be evaluated for accurate approximation. Although the value of K determines the accuracy of the calculation of integral, we do not discuss the relation of K and the evaluation of integral as Butter and Moffitt (1982) did. Several evaluations of the integral using seven periods of arbitrary values of data and coefficients on two right-hand-side variables shows that the value of K is chosen to be 21 is highly accurate. We repeat the maximization of (3.33) for 300 hundreds. We make the notations in the simulation as follows:

1. θ^* means the conditional maximum likelihood estimators.

2. $\hat{\theta} = \frac{1}{300}\sum_{j=1}^{300}\theta_j^*$.

3. θ means true value of parameter, where $\theta = (\rho, \alpha_0, \alpha_1, \gamma, \sigma_a, \sigma_\varepsilon) = (\rho, 0.2, 0.4, \gamma, \sqrt{1.2}, \sqrt{2.4})$.

In each repetition, we proceed with the hypothesis $H_0 : \rho = \rho_0$, where ρ_0 is 0, 0.25, 0.5, 0.75, 0.9, 0.95 when the true value of ρ is 0, 0.25, 0.5, 0.75, 0.9 for each value of ρ_0 with keeping the true values of the other parameters unchanged, (α_0, α_1, γ, σ_a, σ_ε) = (0.2, 0.4, γ, $\sqrt{1.2}$, $\sqrt{2.4}$) where γ is 0 or 0.1. For each hypothesis test for estimates, we calculate the numbers of occurrence that greater than 300 $\times$ 0.01, 0.05 and 0.1 in respective and the result is divided by 300. In other words, we examine the p-value of each estimates under the hypothesis test. Table 3.1 reports the results for true value of ρ=0, ..., 0.9, and γ= 0 and 0.1. The true value of (α_0, α_1, σ_a, σ_ε) is always set to be (0.2, 0.4, $\sqrt{1.2}$, $\sqrt{2.4}$). Table 3.2 reports the p-value when testing value ρ_0 is different from the true value of ρ. To examine the non-normality, Table 3.3 and Table 3.4 report the results of the same test as the previous while the ε_{it} comes from t-distribution with freedom 6 and the c_i from t-distribution with freedom 10.

In case where γ is zero, the performance of the model in Chapter 3 is close to the model in the absence of state dependence in Chapter 2. For example, given $\theta_0 = (0, 0.2, 0.4, \sqrt{2.4}, \sqrt{1.2})$ in Table 2.1-(i), the average of $\hat{\rho}$ is 5×10^{-4} and its p-value is 0.0108, 0.0692, 0.1141 respectively at the corresponding sizes, 0.01, 0.05, 0.1.

Under the same data set in which the parameters is the same as the model in Chapter 2, the average of $\hat{\rho}$ is 2.5×10^{-3} and the p-values are 0.0133, 0.07, 0.1 respectively at the corresponding sizes, 0.01, 0.05, 0.1 in Table 3.1-(i). To check the hypothesis H: $\rho = \rho_0$ when the true value of ρ is zero, Table 3.2-(i) and Table 2.2-(i) show that given the power = 0.01, the two sets of p-value are { 0.0133, 0.1, 0.5200, 0.9133, 0.9933, 1.000,...}, { 0.0108, 0.1167, 0.5400, 0.9050, 0.9925,...} in respective when ρ_0 is $\{0, 0.05, 0.1, \ldots\}$ at the step of 0.05. Although the bias of $\bar{\hat{\rho}}$ is adequately larger, the results gives the numerical evidence that both of models have close power of rejecting wrong when ρ is close to zero.

To check the other extreme case where the value of ρ is getting close to 1, for example $\rho = 0.9$, Table 3.1-(ii) and Table 2.1-(iv) show that $\bar{\hat{\rho}}$ is 0.9004, 0.9018 in respective and the corresponding p-value 0.01, 0.0433, 0.09 and 0.01, 0.0458, 0.0967 in respective at the power of 0.01, 0.05, 0.1 when ρ is 0.5. To check the hypothesis H: $\rho = \rho_0$ when the true value of ρ is 0.9, Table 3.2-(ii) and Table 2.1-(iv) show that given the power = 0.01, the two sets of p-value are { ..., 1.000, 0.9800, 0.5967, 0.0633, 0.0100, 0.1700, 0.5967 } and { ..., 1.000, 0.9850, 0.6558, 0.0908, 0.0100, 0.1658, 0.5975 } in respective when ρ_0 is {..., 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1 } at the step of 0.05.

In case where ρ is not near either 0 or 1, say $\rho = 0.5$, Table 3.1- (i) and Table 2.1- (iii) show $\bar{\hat{\rho}}$ are 0.5004, 0.5021 and the relevant p-value is 0.0167, 0.0333, 0.09 and 0.0100, 0.0500, 0.1050 at the power of 0.01, 0.05, 0.1. To check the hypothesis H: $\rho = \rho_0$ when the true value of ρ is 0.5, Table 3.2-(i) and Table 2.2-(iii) show that given the power = 0.01, the set of p-value are { ..., 0.04, 0.0167, 0.1000, 0.3733, 0.7300, 0.9300,... } and { ...,0.0575, 0.7442, 0.9317, 0.9833, 0.9975,... } when ρ_0 is {...,0.45, 0.5, 0.55,0.6, 0.65, 0.7,... } at the step of 0.05. The numerical evidence shows that the model considering the state dependence have good performance even in the true model in the absence of state dependence.

Comparing Table 3.3 and Table 3.4 with Table 2.3 and Table 2.4 the non-normality cases corresponding to the previous ones have good performance. For example, when ρ is 0.5, Table 3.3- (i) and Table 2.3-(iii) show $\bar{\hat{\rho}}$ are 0.4982, 0.4999 in respective and the relevant p-values are 0.0033, 0.0633, 0.1033 and 0.01, 0.0558, 0.1017 in respective at the power of 0.01, 0.05, 0.1. To check the hypothesis H: $\rho = \rho_0$ when the true value of ρ is 0.5, Table 3.4-(i) and Table 2.3-(iii) show that at the power of 0.01, the two sets of p-values are { ..., 0.07, 0.0033, 0.1433, 0.5233, 0.8933, 0.9833,... } and {..., 0.0792, 0.0100, 0.1183, 0.5217, 0.8842, 0.9842,... } in respective when ρ_0 is {...,0.45, 0.5, 0.55,0.6, 0.65, 0.7,... } at the step of 0.05 .

The numerical evidence support that the application of CMLE into model (3.1) have a good performance even when the true model is in the absence of state dependence.

3.3.2 Model With Strictly Exogenous Variable

In this section, the true model is as follows:

$$\begin{aligned} y_{it} = \rho\ y_{i,t-1} + 0.15\ x_{it} + a_i + \gamma\ [a_i\ y_{i,t-1}] + \varepsilon_{it}, \\ i = 1, \ldots, 250, \quad t = 1, \ldots, 5, \end{aligned} \tag{3.34}$$

where $a_i = 0.2 + 0.4\ y_{i0} + 0.35\ \overline{x_i} + c_i$. The true value of ρ = 0, 0.25, 0.5, 0.75, 0.9, 0.95, and the true value of γ = 0 and 0.1. y_{i0} and x_{it} is generated from N(0,1). We generate ε_{it} by two methods, one is N(0,2.4) and the other is t-distribution with freedom 6. c_i is also generated from N(0,1.2) and t-distribution with freedom 10. By subtracting the constant from the (3.29) and rewrite it as follows:

$$l(y_i, x_i; \theta) \ = \log \int_{-\infty}^{\infty} g(a)\ \exp(-(\frac{a - \mu_a(y_{i0}, x_i)}{\sqrt{2}\sigma_a})^2)\ d\,a, \tag{3.35}$$

where

$$g(a_i) = \left(\frac{1}{\sqrt{\sigma_\varepsilon^2}}\right)^T \left[\prod_{t=1}^{T} \exp\left(\frac{-1}{2\sigma_\varepsilon}(y_{it} - (\rho\ y_{i,t-1} + \beta\ x_{it} + a_i + \gamma\ (a_i y_{i,t-1}))\right)\right]$$

and $a_i = \alpha_0 + \ \alpha_1\ y_{i0} + \alpha_2\ \overline{x_i}$.

Let $z_i = \dfrac{a_i - \mu_a(y_{i0}, \overline{x}_i)}{\sqrt{2}\sigma_a}$. Equation (3.35) can be transformed into the form of

Hermite integral formula as follows:

$$l(y_i, x_i; \theta) \;=\; \log \int_{-\infty}^{\infty} g(z)\, e^{-z^2} \mathrm{d}\, z, \tag{3.36}$$

where

$$g(z_i) = \left(\frac{1}{\sqrt{\sigma_\varepsilon^2}}\right)^T \left[\prod_{t=1}^T \exp\left(\frac{-1}{2\sigma_\varepsilon}(y_{it} - (\rho\, y_{i,t-1} + \beta\, x_{it} + (\sqrt{2}\sigma_a z_i + \mu_a(y_{i0}, \overline{x}_i))+ \right.\right.$$
$$\left.\left. \gamma\, (\sqrt{2}\sigma_a z_i + \mu_a(y_{i0}, \overline{x}_i))))\right)\right].$$

By maximizing $\sum_{i=1}^N \log\{\sum_{j=1}^K w_j g(z_j)\}$, we proceed with the same procedure as that of the previous section. The results are reported from Table 3.5 - 3.8.

Similarly, in each repetition, we proceed with the hypothesis $H_0 : \rho = \rho_0$, where ρ_0 is 0, 0.25, 0.5, 0.75, 0.9, 0.95 when the true value of ρ is 0, 0.25, 0.5, 0.75, 0.9 for each value of ρ_0 with keeping the true values of the other parameters unchanged, $(\alpha_0,\ \alpha_1,\ \gamma,\ \sigma_a,\ \sigma_\varepsilon) = (0.2,\ 0.4,\ \gamma,\ \sqrt{1.2},\ \sqrt{2.4})$ where γ is 0 or 0.1. For each hypothesis test for estimates, we calculate the numbers of occurrence that greater than $300 \times$ 0.01, 0.05 and 0.1 in respective and the result is divided by 300.

We make the comparisons between Table 3.5 - 3.8 and Table 2.6 - 2.10 to show that including the exogenous variables the performance of CMLE for the model with sate dependence when the true model is in the absence of state dependence. In first case where γ is zero, the performance of the model in Chapter 3 has the performance similar to the model in the absence of state dependence in Chapter 2. For example, given $\theta_0 = (0, 0.15, 0.2, 0.4, \sqrt{2.4}, \sqrt{1.2})$ in Table 2.6-(i), $\bar{\hat{\rho}} = 1.7 \times 10^{-3}$ and its p-value is 0.0100, 0.0583, 0.1058, respectively at the corresponding power, 0.01, 0.05, 0.1.

Under the same data set in which the parameters is the same as the model in Chapter 2, $\bar{\hat{\rho}} = 2.2 \times 10^{-3}$ and the p-value is 0.0067, 0.0400, 0.0867 respectively at the corresponding power, 0.01, 0.05, 0.1 in Table 3.5. To check the hypothesis H:

$\rho = \rho_0$ when the true value of ρ is zero, Table 3.6 and Table 2.7-(i) show that given the power = 0.01, the two sets of p-value are { 0.0067, 0.0933, 0.5033, 0.8833, 0.9933, 1.000,...}, { 0.0100, 0.1083, 0.5283, 0.9000, 0.9892.1.000,...} in respective when ρ_0 is { 0, 0.05, 0.1, 0.15, 0.20, 0.25,...} at the step of 0.05. We can check the cases where ρ is 0.9 or 0.5 by the same method used by model (3.30) to compare the current models and the corresponding model in the absence of state dependence in Chapter 2. (see Table 3.5 and 3.6 , and Table 2.6- (iii), (iv) and Table 2.7- (iii), (iv).) In the non-normality cases where γ is zero, we can see the performance in comparison of Table 3.8 and Table 3.9 with Table 2.8- (i), (iii), (iv) and 2.9- (i), (iii), (iv).

With regard to the inclusion of the exogenous variables or not, the performance of the CMLE for models allowing for the interaction between the unobserved effect and the lagged dependence is very well even in the true model which is in the absence of effect of state dependence.

3.4 Empirical example

In this section, we use the data from Vella and Verbeek (1998) to study the conditional maximum likelihood estimator in estimating the AR(1) model in which the unobserved effects interact with the past dependent variable. These data are for young males taken from the National Longitudinal Survey (Youth Sample) for the period 1980-87. As in Chapter 2, we estimate a dynamic log wage equation. We consider the data of log hour wage and the status of labor union. Each of the 545 men in the sample worked in every year from 1980 through 1987. We begin with a single dependent variable, $lnwage_{it}$, to see what is the response of the current wage rate change into the past one in consideration of individual heterogeneity.

It seems reasonable that the amount of state dependence could depend on unobserved heterogeneity. We allow the interaction between the unobserved heterogeneity and the lagged wage rate to account for the heterogenous autoregressive root $(\rho+\gamma a_i)$. An interesting parameter is the average effect, $\vartheta \equiv \rho + \gamma\mu_a$.

We parameterize the model in two ways. We allow for the interaction between the unobserved heterogeneity and the lagged log hour wage as well as the unexplained heterogeneity, a_i is assumed to be $E(a_i|\, lnwage_{i0}, \varepsilon_{it}) = \alpha_0 + \alpha_1 lnwage_{i0} + c_i$ for all i and t. The first case is set up as follows:

$$\begin{aligned} lnwage_{it} = \rho\; lnwage_{i,t-1} +\; a_i + \gamma\; a_i\, lnwage_{i,t-1} + \varepsilon_{it}, \\ i = 1, \ldots, 545, \quad t = 1, \ldots, 7, \end{aligned} \tag{3.37}$$

where $a_i = \alpha_0 + \alpha_1\; lnwage_{i0} +\; c_i, \quad i = 1\ldots, 545.$

In order to obtain a valid standard error for the estimated average effect, rearrange equation (3.37) as follows

$$\begin{aligned} lnwage_{it} = \vartheta\; lnwage_{i,t-1} +\; a_i + \gamma[a_i - \mu_a]\, lnwage_{i,t-1} + \varepsilon_{it}, \\ i = 1, \ldots, 545, \quad t = 1, \ldots, 7, \end{aligned} \tag{3.38}$$

where $\rho = \vartheta - \gamma\,\mu_a$. Models (3.37) and (3.38) are the same model, but formulation (3.38) is convenient because ϑ is the average autoregressive coefficient across population of unobserved heterogeneity. The third case considers adding time dummy variables into equation (3.37) as follows:

$$\begin{aligned} lnwage_{it} = \rho\; lnwage_{i,t-1} + \delta_t\; d_t +\; a_i + \gamma\; a_i\, lnwage_{i,t-1} + \varepsilon_{it}, \\ i = 1, \ldots, 545, \quad t = 1, \ldots, 7, \end{aligned} \tag{3.39}$$

Similar to the manipulation of model (3.37) and (3.38), the amount of state dependence through the average effect of unobserved heterogeneity, $\gamma\,\mu_a\, lnwage_{i,t-1}$ is introduced into equation (3.39) and expressed as follows.

$$\begin{aligned} lnwage_{it} = \vartheta\; lnwage_{i,t-1} + \delta_t\; d_t +\; a_i + \gamma[a_i - \mu_a]\, lnwage_{i,t-1} + \varepsilon_{it}, \\ i = 1, \ldots, 545, \quad t = 1, \ldots, 7, \end{aligned} \tag{3.40}$$

In table 3.9 to 3.14, α_0 and α_1 is significantly greater than zero and $\hat{a}_i$ has positive value. The estimated amount of unobserved heterogeneity of a worker is $\hat{\alpha_0} + \hat{\alpha_1} y_{i0}$ and thus the individual estimated amount of autoregressive coefficient is measured by $\hat{\rho} + \hat{\gamma}\ \hat{a}_i$. That is the response of the current log hourly wage rate into the lagged log hourly wage rate is varied with individual worker, where the estimated size of difference is measured by $\gamma\, \hat{a}_i$. We are more concerned with the average effect, $\rho + \gamma E(a_i)$, and its corresponding estimated value is $\hat{\rho} + \hat{\gamma}\hat{\mu}_a$. In equation (3.37), we calculate the value of $\hat{\rho} + \hat{\gamma} \times \hat{\mu_a} \simeq 0.376$. $\hat{\mu_a}$ is calculated by $\hat{\alpha_0} + \hat{\alpha_1}\ \overline{lnwage}_0$, where $\overline{lnwage}_0 = \frac{1}{545}\sum_{i=1}^{545} lnwage_{i0}$. The estimated amount of average effect can be obtained from estimating the ϑ in equation (3.38). Table 3.10 shows $\hat{\vartheta}$ is 0.3755 significantly greater than zero. The same manipulation on equations (3.39) and (3.40) give evidence that the estimated average effect is about 0.215, (see Table 3.12). Table 3.12 shows $\hat{\vartheta}$ is significantly greater than zero. Empirically, we often consider the model with exogenous variables, for example the labor union membership, once we control for state dependence and unobserved heterogeneity, does union membership matter? Under the assumption that the labor union membership is strictly exogenous we add it to the basic equation. Because we assume that labor union membership is strictly exogenous variable, once the past log hour wage rate, the current log hourly wage rate is not affected by past or future labor union membership. The assumption that reasonable because in general employers might just see if the employees have the labor union membership at present in determining the level of hour wage. There could be feedback from wage innovation to future union membership, although this is possibly small.

We write the model with strictly exogenous variable as follows:

$$lnwage_{it} = \rho\ lnwage_{i,t-1} + \beta\ union_{it} + \ a_i + \gamma\ a_i\, lnwage_{i,t-1} + \varepsilon_{it}, \qquad (3.41)$$
$$i = 1, \ldots, 545, \quad t = 1, \ldots, 7,$$

and we assume

$$a_i = \alpha_0 + \alpha_1\ lnwage_{i0} + \alpha_2\ \overline{union}_i + \ c_i, \quad i = 1 \ldots, 545,$$

and its corresponding average effect, $\rho + \gamma\mu_a$, expression model is

$$lnwage_{it} = \vartheta\, lnwage_{i,t-1} + \beta\ union_{it} + \ a_i + \gamma\ [a_i - \mu_a]\, lnwage_{i,t-1} + \varepsilon_{it}, \qquad (3.42)$$
$$i = 1, \ldots, 545, \quad t = 1, \ldots, 7,$$

where $a_i = \alpha_0 + \alpha_1\ lnwage_{i0} + \alpha_2\ \overline{union}_i + \ c_i$ i=1,. . .,545 and $\mu_a = \alpha_0 + \alpha_1\ \overline{lnwage}_0 + \alpha_2\ \overline{union}$, where $\overline{union}_i$ is the fraction of employment membership. That is the length of keeping the labor union membership more or less reflects the individual preference of a worker. Empirically, the smaller is the ratio, the lower is willingness to keep membership. As often as a worker with higher ability less intend to keep union membership. We report the CMLE estimates of the model (3.41) in Table 3.13. The estimates of $\hat{\beta}$ is significant and $\hat{\alpha_2}$ is not significantly different from zero. Except for the $\hat{\gamma}$, the other estimates is very close to those of model (3.37). The estimated amount of individual autoregressive coefficient is $\hat{\rho}\ +\ \gamma\ \hat{a}_i$, where $\hat{a}_i = \alpha_0 + \alpha_1 y_{i0} + \alpha_2 \overline{union}_i$. Table 3.13 shows that $(\hat{\rho}, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\gamma})$ is (-0.4771, 0.0507, 0.943, 0.082, 0.0244, 0.7757).

The same manipulation as the model without exogenous variable the estimated amount of average effect is measured by $\hat{\rho} + \hat{\gamma} \times \hat{\mu_a} \simeq 0.3476$. $\hat{\mu}_a$ is calculated by $\hat{\alpha_0}\ +\ \hat{\alpha_1}\ \overline{lnwage}_0 + \alpha_2 \overline{union}$, where $\overline{lnwage}_0\ =\ \frac{1}{545}\sum_{i=1}^{545} lnwage_{i0}$ and $\overline{union}\ = \frac{1}{545}\sum_{i=1}^{545} \overline{union}_i$ which means the average length of period for keeping the membership for the workers we observed. Model (3.39) gives $\hat{\vartheta} = 0.3480$, (see Table 3.14). The log hourly wage rate is more or less influenced by the labor union membership. However, the estimates of $\hat{\alpha_2}$ is not significantly different from zero in the log hourly wage equation. The linear relationship between labor union membership on the unobserved heterogeneity is small. Similar to equation (3.38), we report model (3.42) in Table 3.14.

We report the empirical examples in Table 3.9 - 3.14. At last we report the range of estimate of the response to the future $(\rho+\gamma\,\mu_a)$ by measuring $(\hat{\rho}+\hat{\gamma}\,(\hat{\mu}_a \pm std.(a_i)))$. The results for the range of models (3.37) and (3.38) are 0.376 $\pm$ 0.829(0.090). The range of models (3.39) and (3.40) are 0.215 $\pm$ 0.134(0.219).

3.5 Conclusion

In this chapter we apply the CMLE in estimating a panel data model where unobserved heterogeneity interacts with a lagged dependent variable. The IV estimator for the coefficient of $y_{i,t-1}$ is inconsistent even for the average effect. In other words, the existing approach can not estimate the amount of average state dependence. Our recommended approach for the model is flexible. Firstly, we just model the distribution of the unobserved heterogeneity, $D(a_i|x_i, y_{i0})$, and then construct the density of $(y_{i1}, \ldots, y_{iT})$ given (x_i, y_{i0}). To specify a conditional parametric density function of a_i in which the conditional mean and variance are flexible function of y_{i0} and x_i. We can easily define a log-likelihood function conditional on (x_i, y_{i0}). It is crucial point that the conditional maximum likelihood function is valid no matter what we condition on. Furthermore, the conditional maximum likelihood function is consistent, $\sqrt{N}$-asymptotically normal, under standard regularity conditions. Secondly, we can easily estimate ρ and γ as well as $(\rho\ +\gamma\,\mu_a)$.

The idea of specifying a distribution for the unobserved effects given the initial condition to construct CMLE means we can easily estimate the average partial effect across a_i. If x_i is not strictly exogenous, we can apply the suggestion of Wooldridge (2000a) as follows. We parameterize $g(x_t|X_{t-1}, z_t, a; \lambda_0)$, where z_t is strictly exogenous and build up the joint density of (y_t, x_t) given $(z_T, Y_{t-1}, X_{t-1}, a)$ and then apply the same procedure as discussed previously to set up a log-likelihood function. We can use numerical methods to solve out the CMLE.

In the empirical example of hourly log wage rate, the interaction of lagged wage and the unobserved heterogeneity is significant. This restricts the case where we add union status to the model. Union status is marginally significant and is estimated to increase wage by about 5 percent.

We need to add more explanatory variables and apply the more general method allowing for non-strict exogeneity assumption to decrease the degree of the interdependence between the unobserved heterogeneity and the error term. We propose the basic model to illustrate conditional MLE. These models can be easily extended to more complicate case. It is crucial problem that we need to find an appropriate formula to approximate the integral to integrate out the unobserved heterogeneity. The more complicated is the conditional parametric density of the unobserved heterogeneity, the more difficult it is to find a good formula for approximating the necessary integral. The Gauss quadrature method seems to work well, is time-consuming. Future research could focus on simulation methods of estimation, as in Keane(1993).

Table 3.1: $H_0 : \theta = \theta_0$, where ρ =0 $\sim$ 0.9

$$\theta \quad = (\rho, 0.2, 0.4, \gamma, \sqrt{2.4}, \sqrt{1.2})$$
$$\theta_0 \quad = (\rho_0, 0.2, 0.4, \gamma_0, \sqrt{2.4}, \sqrt{1.2})$$

$P\backslash\hat{\theta}$	2.5×10^{-3}	0.2022	0.3966	1.1×10^{-3}	1.5502	1.0950	(ρ,γ)
	3×10^{-3}	0.2025	0.3964	0.1022	1.5503	1.0804	
0.01	0.0133	0.0067	0.0100	0.0067	0.0167	0.0067	(0,0)
	0.0167	0.0033	0.0133	0.0067	0.0167	0.0100	(0,0.1)
0.05	0.0700	0.0367	0.0433	0.0367	0.0633	0.0533	(0,0)
	0.0600	0.0500	0.0467	0.0400	0.0567	0.0500	(0,0.1)
0.10	0.1000	0.0800	0.0867	0.0933	0.1000	0.0900	(0,0)
	0.1033	0.0867	0.0800	0.1033	0.1100	0.0867	(0,0.1)
$P\backslash\hat{\theta}$	0.2515	0.2017	0.3970	1.3×10^{-3}	1.5502	1.0804	(ρ,γ)
	0.2519	0.2022	0.3967	0.1025	1.5502	1.0811	
0.01	0.0133	0.0067	0.0100	0.0033	0.0200	0.0033	(0.25,0)
	0.0100	0.0033	0.0067	0.0033	0.0167	0.0000	(0.25,0.1)
0.05	0.0600	0.0467	0.0467	0.0433	0.0567	0.0433	(0.25,0)
	0.0433	0.0533	0.0533	0.0433	0.0533	0.0433	(0.25,0.1)
0.10	0.0933	0.0867	0.0833	0.1033	0.1100	0.0867	(0.25,0)
	0.0967	0.0933	0.1000	0.1000	0.1067	0.0833	(0.25,0.1)
$P\backslash\hat{\theta}$	0.5004	0.2014	0.3978	1.3×10^{-3}	1.5500	1.0818	(ρ,γ)
	0.5004	0.2019	0.3979	0.1028	1.5496	1.0838	
0.01	0.0167	0.0033	0.0100	0.0033	0.0167	0.0033	(0.5,0)
	0.0133	0.0067	0.0200	0.0067	0.0133	0.0033	(0.5,0.1)
0.05	0.0333	0.0467	0.0500	0.0467	0.0667	0.0300	(0.5,0)
	0.0333	0.0567	0.0600	0.0533	0.0533	0.0400	(0.5,0.1)
0.10	0.0900	0.0833	0.0900	0.1000	0.0967	0.0767	(0.5,0)
	0.0833	0.0967	0.0967	0.0933	0.1000	0.0800	(0.5,0.1)

Normality

Rrepetitions $= 300, \hat{\theta} = \frac{1}{300}\sum_{j=1}^{300}\theta_j^*, \sqrt{2.4} \simeq 1.5492, \sqrt{1.2} \simeq 1.0954$

The content of bracket is $(\rho_0\ ,\ \gamma_0)$

(i)

P\ $\hat{\theta}$							
P\ $\hat{\theta}$	0.7504	0.2015	0.3984	6×10^{-4}	1.5500	1.0815	(ρ,γ)
	0.7484	0.2010	0.3998	0.1025	1.5482	1.0893	
0.01	0.0133	0.0067	0.0167	0.0067	0.0200	0.0067	(0.75,0)
	0.0133	0.0100	0.0267	0.0067	0.0167	0.0067	(0.75,0.1)
0.05	0.0500	0.0533	0.0567	0.0333	0.0633	0.0200	(0.75,0)
	0.0467	0.0600	0.0733	0.0533	0.0600	0.0200	(0.75,0.1)
0.10	0.1000	0.0900	0.0900	0.1100	0.1067	0.0567	(0.75,0)
	0.1000	0.1000	0.0967	0.1033	0.0900	0.0633	(0.75,0.1)
P\ $\hat{\theta}$	0.9004	0.2016	0.3987	2×10^{-4}	1.5498	1.0822	(ρ,γ)
	0.8980	0.2001	0.3993	0.1019	1.5472	1.0917	
0.01	0.0100	0.0067	0.0167	0.0067	0.0200	0.0033	(0.9,0)
	0.0067	0.0067	0.09300	0.0067	0.0167	0.0033	(0.9,0.1)
0.05	0.0433	0.0533	0.0567	0.0367	0.0700	0.0267	(0.9,0)
	0.0733	0.0600	0.0800	0.0500	0.0567	0.0200	(0.9,0.1)
0.10	0.0900	0.1000	0.1067	0.1033	0.1133	0.0500	(0.9,0)
	0.1100	0.0967	0.1233	0.1267	0.0967	0.0733	(0.9,0.1)

Rrepetitions $= 300, \hat{\theta} = \frac{1}{300}\sum_{j=1}^{300}\theta_j^*, \sqrt{2.4} \simeq 1.5492, \sqrt{1.2} \simeq 1.0954$

The content of bracket is $(\rho_0\ ,\ \gamma_0)$

(ii)

Table 3.2: $H_0 : \hat{\theta} = \theta_0$, where ρ =0 ~ 0.5

$$\begin{aligned} \theta &= (\rho, 0.2, 0.4, \gamma, \sqrt{2.4}, \sqrt{1.2}) \\ \theta_0 &= (\rho_0, 0.2, 0.4, \gamma_0, \sqrt{2.4}, \sqrt{1.2}) \end{aligned}$$

P\ $\rho_0 \overset{=}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	(ρ,γ)
0.01	0.0133	0.1000	0.5200	0.9133	0.9933	1.0000	1.0000	1.0000	(0,0)
	0.0167	0.0833	0.5067	0.9033	0.9900	1.0000	1.0000	1.0000	(0,0.1)
0.05	0.0700	0.2567	0.7433	0.9533	0.9967	1.0000	1.0000	1.0000	(0,0)
	0.0600	0.2433	0.7600	0.9567	1.0000	1.0000	1.0000	1.0000	(0,0.1)
0.10	0.1000	0.3633	0.8267	0.9833	1.0000	1.0000	1.0000	1.0000	(0,0)
	0.1033	0.3700	0.8400	0.9833	1.0000	1.0000	1.0000	1.0000	(0,0.1)
P\ $\rho_0 \overset{=}{\rightarrow}$	0	0.1	0.15	0.2	0.25	0.3	0.35	0.4	(ρ,γ)
0.01	1.0000	0.8900	0.4567	0.0767	0.0133	0.0900	0.4167	0.8333	(0.25,0)
	1.0000	0.9133	0.4500	0.0867	0.0100	0.0767	0.4400	0.8500	(0.25,0.1)
0.05	1.0000	0.9967	0.6667	0.2167	0.0600	0.2400	0.6767	0.9267	(0.25,0)
	1.0000	0.9800	0.7033	0.2133	0.0433	0.2300	0.6800	0.9333	(0.25,0.1)
0.10	1.0000	0.9800	0.8033	0.3067	0.0933	0.3433	0.7767	0.9533	(0.25,0)
	1.0000	0.9867	0.8100	0.3233	0.0967	0.3367	0.7733	0.9600	(0.25,0.1)
P\ $\rho_0 \overset{=}{\rightarrow}$	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	(ρ,γ)
0.01	0.3267	0.0400	0.0167	0.1000	0.3733	0.7300	0.9300	0.9967	(0.5,0)
	0.3900	0.0500	0.0133	0.1167	0.4300	0.8100	0.9767	0.9967	(0.5,0.1)
0.05	0.6300	0.1533	0.0333	0.2233	0.5967	0.8833	0.9867	0.9967	(0.5,0)
	0.6867	0.2033	0.0333	0.2233	0.6567	0.9167	0.9967	0.9967	(0.5,0.1)
0.10	0.7233	0.2800	0.0900	0.3233	0.7067	0.9267	0.9967	0.9967	(0.5,0)
	0.7867	0.3200	0.0833	0.3267	0.7267	0.9600	0.9967	1.0000	(0.5,0.1)

Normality

Rrepetitions $= 300, \hat{\theta} = \frac{1}{300}\sum_{j=1}^{300}\theta_j^*, \sqrt{2.4} \simeq 1.5492, \sqrt{1.2} \simeq 1.0954$

The content of bracket is $(\rho_0\ ,\ \gamma_0)$

(i)

P\ $\overset{\rho_0=}{\rightarrow}$	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	(ρ,γ)
0.01	0.8500	0.3533	0.0267	0.0133	0.1300	0.4500	0.7700	0.9433	(0.75,0)
	0.9233	0.4767	0.0633	0.0133	0.1400	0.5433	0.8733	0.9833	(0.75,0.1)
0.05	0.9567	0.6667	0.1633	0.0500	0.2667	0.6233	0.8933	0.9767	(0.75,0)
	0.9833	0.7567	0.2300	0.0467	0.3233	0.7433	0.9533	1.0000	(0.75,0.1)
0.10	0.9867	0.7733	0.2867	0.1000	0.3767	0.7233	0.9433	0.9900	(0.75,0)
	0.9900	0.8467	0.3500	0.1000	0.4267	0.8267	0.9700	1.0000	(0.75,0.1)
P\ $\overset{\rho_0=}{\rightarrow}$	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	(ρ,γ)
0.01	1.0000	1.0000	0.9800	0.5967	0.0633	0.0100	0.1700	0.5967	(0.9,0)
	1.0000	1.0000	1.0000	0.7467	0.1367	0.0067	0.2200	0.7500	(0.9,0.25)
0.05	1.0000	1.0000	0.9967	0.8200	0.2600	0.0433	0.3500	0.7667	(0.9,0)
	1.0000	1.0000	1.0000	0.9033	0.3400	0.0733	0.4433	0.8900	(0.9,0.25)
0.10	1.0000	1.0000	1.0000	0.8900	0.3800	0.0900	0.4700	0.8367	(0.9,0)
	1.0000	1.0000	1.0000	0.9367	0.4400	0.1100	0.5267	0.9200	(0.9,0.25)

$Rrepetitions = 300, \hat{\theta} = \frac{1}{300}\sum_{j=1}^{300}\theta_j^*, \sqrt{2.4} \simeq 1.5492, \sqrt{1.2} \simeq 1.0954$

The content of bracket is $(\rho_0\ ,\ \gamma_0)$

(ii)

Table 3.3: $H_0 : \theta = \theta_0$, where ρ =0 ~ 0.9

$$\theta \quad = (\rho, 0.2, 0.4, \gamma, \sqrt{1.5}, \sqrt{1.25})$$
$$\theta_0 \quad = (\rho_0, 0.2, 0.4, \gamma_0, \sqrt{1.5}, \sqrt{1.25})$$

$P\backslash\hat{\theta}$							
	8×10^{-4}	0.2015	0.3984	5×10^{-4}	1.2230	1.0985	(0,0)
	-1.35×10^{-2}	0.1884	0.3870	0.0947	1.2245	1.0992	(0,0.1)
0.01	0.0067	0.0133	0.0167	0.0333	0.0700	0.0133	(0,0)
	0.0233	0.0133	0.0100	0.0433	0.0667	0.0133	(0,0.1)
0.05	0.0533	0.0533	0.0467	0.0933	0.1200	0.0800	(0,0)
	0.0700	0.0600	0.0500	0.1100	0.1133	0.0767	(0,0.1)
0.10	0.1067	0.0967	0.1100	0.1567	0.2000	0.1500	(0,0)
	0.1167	0.0967	0.1067	0.1567	0.2000	0.1467	(0,0.1)
$P\backslash\hat{\theta}$	0.2498	0.2020	0.3988	2×10^{-4}	1.2228	1.0996	(0.25,0)
	0.2364	0.1841	0.3863	0.0950	1.2244	1.0991	(0.25,0.1)
0.01	0.0067	0.0167	0.0133	0.0333	0.0567	0.0167	(0.25,0)
	0.0200	0.0133	0.0133	0.0400	0.0533	0.0167	(0.25,0.1)
0.05	0.0467	0.0567	0.0433	0.0900	0.1167	0.0667	(0.25,0)
	0.0567	0.0533	0.0567	0.1167	0.1200	0.0700	(0.25,0.1)
0.10	0.1000	0.0967	0.0900	0.1567	0.1933	0.1300	(0.25,0)
	0.1067	0.0933	0.1067	0.1800	0.2067	0.1133	(0.25,0.1)
$P\backslash\hat{\theta}$	0.4982	0.2027	0.4033	-0.0000	1.2222	1.1030	(0.5,0)
	0.4846	0.1794	0.3870	0.0944	1.2236	1.1036	(0.5,0.1)
0.01	0.0033	0.0100	0.0100	0.0433	0.0467	0.0233	(0.5,0)
	0.0233	0.0100	0.0167	0.0367	0.0500	0.0200	(0.5,0.1)
0.05	0.0633	0.0500	0.0400	0.0967	0.1333	0.0667	(0.5,0)
	0.0733	0.0533	0.0500	0.1233	0.1267	0.0667	(0.5,0.1)
0.10	0.1033	0.0933	0.0900	0.1733	0.2033	0.1000	(0.5,0)
	0.1367	0.0867	0.1000	0.2033	0.2067	0.1233	(0.5,0.1)

Non-normality

Rrepetitions $= 300, \hat{\theta} = \frac{1}{300}\sum_{j=1}^{300}\theta_j^*, \sqrt{1.5} \simeq 1.2247, \sqrt{1.2} \simeq 1.1180$

The content of bracket is $(\rho_0\ ,\ \gamma_0)$

(i)

P\ $\hat{\theta}$	0.7469	0.2032	0.4028	-2×10^{-4}	1.2213	1.1077	(0.75,0)
	0.7338	0.1720	0.3855	0.0925	1.2217	1.1126	(0.75,0.1)
0.01	0.0100	0.0100	0.0100	0.0333	0.0500	0.0200	(0.75,0)
	0.0333	0.0267	0.0167	0.08670	0.0600	0.0267	(0.75,0.1)
0.05	0.0567	0.0533	0.0433	0.1133	0.1433	0.0567	(0.75,0)
	0.1000	0.0467	0.0600	0.1800	0.1467	0.0567	(0.75,0.1)
0.10	0.1233	0.0833	0.0900	0.1600	0.2033	0.1100	(0.75,0)
	0.1600	0.1133	0.1133	0.2567	0.2100	0.1267	(0.75,0.1)
P\ $\hat{\theta}$	0.8971	0.2034	0.4039	-2×10^{-4}	1.2210	1.1090	(0.9,0)
	0.8852	0.1707	0.3833	0.0915	1.2224	1.1161	(0.9,0.1)
0.01	0.0133	0.0067	0.0100	0.0433	0.0467	0.0133	(0.9,0)
	0.0420	0.0180	0.0260	0.1340	0.0600	0.0200	(0.9,0.1)
0.05	0.0567	0.0567	0.0533	0.1233	0.1500	0.0700	(0.9,0)
	0.1140	0.0440	0.0860	0.2380	0.1400	0.0700	(0.9,0.1)
0.10	0.1133	0.0833	0.0967	0.1533	0.2167	0.1167	(0.9,0)
	0.2000	0.1180	0.1320	0.3240	0.2180	0.1220	(0.9,0.1)

Rrepetitions $= 300, \hat{\theta} = \frac{1}{300}\sum_{j=1}^{300}\theta_j^*, \sqrt{1.5} \simeq 1.2247, \sqrt{1.2} \simeq 1.1180$

The content of bracket is $(\rho_0 \ , \ \gamma_0)$

(ii)

Table 3.4: $H_0 : \theta = \theta_0$, where ρ =0 ~ 0.9

$$\theta = (\rho, 0.2, 0.4, \gamma, \sqrt{1.5}, \sqrt{1.25})$$
$$\theta_0 = (\rho_0, 0.2, 0.4, \gamma_0, \sqrt{1.5}, \sqrt{1.25})$$

P\ $\rho_0 =$ →	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	(ρ, γ)
0.01	0.0067	0.1267	0.5767	0.9200	1.0000	1.0000	1.0000	1.0000	(0,0)
	0.0233	0.2400	0.7067	0.9667	1.0000	1.0000	1.0000	1.0000	(0,0.1)
0.05	0.0533	0.2900	0.7733	0.9733	1.0000	1.0000	1.0000	1.0000	(0,0)
	0.0700	0.4433	0.8800	0.9967	1.0000	1.0000	1.0000	1.0000	(0,0.1)
0.10	0.1067	0.4000	0.8533	0.9867	1.0000	1.0000	1.0000	1.0000	(0,0)
	0.1167	0.5600	0.9300	1.0000	1.0000	1.0000	1.0000	1.0000	(0,0.1)
P\ $\rho_0 =$ →	0	0.1	0.15	0.2	0.25	0.3	0.35	0.4	(ρ, γ)
0.01	1.0000	0.9133	0.5133	0.0800	0.0067	0.1267	0.5200	0.8867	(0.25,0)
	1.0000	0.8700	0.3800	0.0367	0.0200	0.2133	0.6467	0.9500	(0.25,0.1)
0.05	1.0000	0.9833	0.7433	0.2300	0.0467	0.2867	0.7233	0.9567	(0.25,0)
	1.0000	0.9633	0.6167	0.1533	0.0567	0.4133	0.8333	0.9867	(0.25,0.1)
0.10	1.0000	0.9933	0.8267	0.3800	0.1000	0.3800	0.8300	0.9700	(0.25,0)
	1.0000	0.9800	0.7333	0.2433	0.1067	0.5200	0.9167	0.9933	(0.25,0.1)
P\ $\rho_0 =$ →	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	(ρ, γ)
0.01	0.4833	0.0700	0.0033	0.1433	0.5233	0.8933	0.9833	0.9967	(0.5,0)
0.01	0.3700	0.0300	0.0233	0.2367	0.6767	0.9533	0.9933	1.0000	(0.5,0.1)
0.05	0.7133	0.2067	0.0633	0.3100	0.7200	0.9533	0.9933	1.0000	(0.5,0)
0.05	0.6033	0.1300	0.0733	0.4500	0.8400	0.9833	1.0000	1.0000	(0.5,0.1)
0.10	0.8267	0.3533	0.1033	0.4100	0.8267	0.9733	0.9967	1.0000	(0.5,0)
0.10	0.7333	0.2233	0.1367	0.5500	0.9033	0.9900	1.0000	1.0000	(0.5,0.1)

Non-normality

Rrepetitions = 300, $\hat{\theta} = \frac{1}{300}\sum_{j=1}^{300}\theta_j^*$, $\sqrt{1.5} \simeq 1.2247$, $\sqrt{1.2} \simeq 1.1180$

The content of bracket is $(\rho_0\ ,\ \gamma_0)$

(ii)

$P\backslash \overset{\rho_0=}{\rightarrow}$	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	(ρ,γ)
0.01	0.9733	0.6467	0.0933	0.0100	0.1900	0.6733	0.9533	0.9900	(0.75,0)
	0.9567	0.5100	0.0433	0.0333	0.3133	0.8000	0.9867	0.9967	(0.75,0.1)
0.05	0.9967	0.8333	0.2767	0.0567	0.3733	0.8400	0.9833	0.9967	(0.75,0)
	0.9933	0.7167	0.1800	0.1000	0.5467	0.9233	0.9933	0.9967	(0.75,0.1)
0.10	1.0000	0.8933	0.3900	0.1233	0.5000	0.8867	0.9900	0.9967	(0.75,0)
	1.0000	0.8100	0.2800	0.1600	0.6300	0.9633	0.9933	1.0000	(0.75,0.1)

$P\backslash \overset{\rho_0=}{\rightarrow}$	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	(ρ,γ)
0.01	1.0000	1.000	1.0000	0.8233	0.1567	0.0133	0.2733	0.8367	(0.9,0)
	1.0000	1.000	0.9900	0.7120	0.0840	0.0420	0.4620	0.9300	(0.9,0.1)
0.05	1.0000	1.0000	1.0000	0.9500	0.3800	0.0567	0.5233	0.9433	(0.9,0)
	1.0000	1.0000	1.0000	0.8820	0.2740	0.1140	0.6520	0.9760	(0.9,0.1)
0.10	1.0000	1.0000	1.0000	0.9800	0.5200	0.1133	0.6067	0.9767	(0.9,0)
	1.0000	1.0000	1.0000	0.9280	0.3920	0.2000	0.7560	0.9840	(0.9,0.1)

$Rrepetitions = 300, \hat{\theta} = \frac{1}{300}\sum_{j=1}^{300}\theta_j^*, \sqrt{1.5} \simeq 1.2247, \sqrt{1.2} \simeq 1.1180$

$The\ content\ of\ bracket\ is\ (\rho_0\ ,\ \gamma_0)$

(ii)

Table 3.5: $H_0 : \theta = \theta_0$, where ρ =0 $\sim$ 0.9

$$\theta \quad = (\rho, 0.15, 0.2, 0.4, 0.35, \gamma, \sqrt{2.4}, \sqrt{1.2})'$$
$$\theta_0 \quad = (\rho_0, 0.15, 0.2, 0.4, 0.35, \gamma_0, \sqrt{2.4}, \sqrt{1.2})$$

P\\$\hat{\theta}$									(ρ, γ)
	0.0022	0.1530	0.2018	0.3956	0.3355	-0.0002	1.5471	1.0828	
	0.0019	0.1529	0.2027	0.3962	0.3362	0.1001	1.5470	1.0841	
0.01	0.0067	0.0100	0.0067	0.0167	0.0100	0.0033	0.0067	0.0067	(0,0)
	0.0133	0.0100	0.0067	0.0100	0.0100	0.0067	0.0067	0.0067	(0,0.1)
0.05	0.0400	0.0367	0.0633	0.0567	0.0400	0.0500	0.0567	0.0433	(0,0)
	0.0567	0.0333	0.0667	0.0567	0.0400	0.0467	0.0533	0.0467	(0,0.1)
0.10	0.0867	0.0867	0.0967	0.1167	0.0933	0.0800	0.1100	0.1133	(0,0)
	0.0933	0.0900	0.0967	0.1233	0.1033	0.0967	0.1100	0.1033	(0,0.1)
P\\$\hat{\theta}$	0.5010	0.1530	0.2024	0.3960	0.3365	-0.0003	1.5473	1.0838	(ρ, γ)
	0.5010	0.1529	0.2028	0.3958	0.3368	0.1009	1.5472	1.0841	
0.01	0.0067	0.0100	0.0100	0.0133	0.0133	0.0033	0.0067	0.0033	(0.5,0)
	0.0167	0.0100	0.0067	0.0100	0.0133	0.0100	0.0067	0.0067	(0.5,0.1)
0.05	0.0533	0.0367	0.0633	0.0533	0.0433	0.0433	0.0400	0.0433	(0.5,0)
	0.0533	0.0267	0.0667	0.0567	0.0533	0.0433	0.0500	0.0467	(0.5,0.1)
0.10	0.1067	0.0867	0.1033	0.1100	0.0900	0.0733	0.0967	0.0867	(0.5,0)
	0.1000	0.0967	0.0967	0.0967	0.0800	0.0867	0.1100	0.1133	(0.5,0.1)
P\\$\hat{\theta}$	0.9009	0.1533	0.2020	0.3966	0.3384	-0.0002	1.5472	1.0846	(ρ, γ)
	0.8997	0.1531	0.2015	0.3997	0.3365	0.1016	1.5463	1.0879	
0.01	0.0133	0.0100	0.0133	0.0067	0.0167	0.0033	0.0100	0.0167	(0.9,0)
	0.0167	0.0067	0.0100	0.0167	0.0200	0.0167	0.0133	0.0133	(0.9,0.1)
0.05	0.0533	0.0400	0.0600	0.0533	0.0467	0.0433	0.0433	0.0500	(0.9,0)
	0.0833	0.0367	0.0600	0.0667	0.0500	0.0667	0.0533	0.0833	(0.9,0.1)
0.10	0.1167	0.0900	0.1033	0.1200	0.0833	0.0867	0.0933	0.1067	(0.9,0)
	0.1400	0.1033	0.1067	0.1233	0.1000	0.1267	0.1133	0.1433	(0.9,0.1)

Normality

$Rrepetitions = 300, \hat{\theta} = \frac{1}{300}\sum_{j=1}^{300}\theta_j^*, \sqrt{1.5} \simeq 1.5492, \sqrt{1.2} \simeq 1.0954$

The content of bracket is $(\rho_0\ ,\ \gamma_0)$

Table 3.6: $H_0 : \theta = \theta_0$, ρ =0 $\sim$ 0.5

$$\theta \quad = (\rho, 0.2, 0.4, \gamma, \sqrt{2.4}, \sqrt{1.2})$$
$$\theta_0 \quad = (\rho_0, 0.2, 0.4, \gamma_0, \sqrt{2.4}, \sqrt{1.2})$$

P\ $\overset{\rho_0=}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	(ρ,γ)
0.01	0.0067	0.0933	0.5033	0.8833	0.9933	1.0000	1.0000	1.0000	(0,0)
	0.0133	0.1133	0.5067	0.9000	1.0000	1.0000	1.0000	1.0000	(0,0.1)
0.05	0.0400	0.2700	0.7167	0.9800	0.9967	1.0000	1.0000	1.0000	(0,0)
	0.0567	0.2667	0.7333	0.9733	1.0000	1.0000	1.0000	1.0000	(0,0.1)
0.10	0.0867	0.3767	0.8000	0.9833	1.0000	1.0000	1.0000	1.0000	(0,0)
	0.0933	0.3567	0.8100	0.9867	1.0000	1.0000	1.0000	1.0000	(0,0.1)

P\ $\overset{\rho_0=}{\rightarrow}$	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	(ρ,γ)
0.01	0.8133	0.4000	0.0400	0.0067	0.1100	0.4000	0.7400	0.9367	(0.5,0)
	0.8467	0.4567	0.0667	0.0167	0.1133	0.4033	0.8033	0.9733	(0.5,0.1)
0.05	0.9367	0.6367	0.2000	0.0533	0.2367	0.5733	0.8767	0.9900	(0.5,0)
	0.9600	0.6667	0.2133	0.0533	0.2333	0.6067	0.9267	0.9933	(0.5,0.1)
0.10	0.9700	0.7333	0.3000	0.1067	0.3167	0.6900	0.9300	0.9933	(0.5,0)
0.9733	0.7767	0.3367	0.1000	0.3567	0.7567	0.9533	0.9967	(0.5,0.1)	

P\ $\overset{\rho_0=}{\rightarrow}$	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	(ρ,γ)
0.01	1.0000	1.0000	1.0000	0.9733	0.5967	0.0767	0.0133	0.1733	(0.9,0)
	1.0000	1.0000	0.9800	0.7333	0.1367	0.0167	0.2033	0.7167	(0.9,0.1)
0.05	1.0000	1.0000	1.0000	0.8200	0.2500	0.0533	0.3300	0.7533	(0.9,0)
	1.0000	1.0000	1.0000	0.8900	0.3367	0.0833	0.3800	0.8467	(0.9,0.1)
0.10	1.0000	1.0000	1.0000	0.8200	0.2500	0.0533	0.3300	0.7533	(0.9,0)
	1.0000	1.0000	1.0000	0.9233	0.4467	0.1400	0.4967	0.9033	(0.9,0.1)

Rrepetitions $= 300, \hat{\theta} = \frac{1}{300}\sum_{j=1}^{300}\theta_j^*$, $\sqrt{2.4} \simeq 1.5492$, $\sqrt{1.2} \simeq 1.0954$
The content of bracket is $(\rho_0 \ , \ \gamma_0)$
Normality

Table 3.7: $H_0 : \theta = \theta_0$, where ρ =0 $\sim$ 0.9

$$\theta = (\rho, 0.15, 0.2, 0.4, 0.35, \gamma, \sqrt{1.5}, \sqrt{1.25})$$
$$\theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \gamma_0, \sqrt{1.5}, \sqrt{1.25})$$

P\\$\hat{\theta}$	0.0004	0.1521	0.2017	0.3986	0.3451	0.0005	1.2223	1.0964	(ρ,γ)
	-0.0032	0.1521	0.2424	0.4141	0.3429	0.0170	1.2228	1.0996	
0.01	0.0100	0.0133	0.0133	0.0167	0.0067	0.0367	0.0700	0.0133	(0,0)
	0.0133	0.0133	0.0200	0.0133	0.0067	0.9533	0.0700	0.0133	(0,0.1)
0.05	0.0600	0.0600	0.0467	0.0433	0.0400	0.0933	0.1200	0.0867	(0,0)
	0.0633	0.0600	0.0733	0.0533	0.0400	0.9767	0.1167	0.0800	(0,0.1)
0.10	0.1067	0.0933	0.1067	0.1100	0.0900	0.1467	0.1900	0.1400	(0,0)
	0.1167	0.0933	0.1500	0.1233	0.0967	0.9900	0.1933	0.1500	(0,0.1)

P\\$\hat{\theta}$	0.4975	0.1519	0.2028	0.4009	0.3472	0.0001	1.2214	1.1015	(ρ,γ)
	0.4973	0.1518	0.1988	0.3997	0.3441	0.1021	1.2210	1.1006	
0.01	0.0133	0.0133	0.0133	0.0133	0.0067	0.0333	0.0550	0.0167	(0.5,0)
	0.0167	0.0133	0.0100	0.0133	0.0100	0.0567	0.0567	0.0167	(0.5,0.1)
0.05	0.0600	0.0633	0.0500	0.0333	0.0400	0.0933	0.1300	0.0533	(0.5,0)
	0.0433	0.0667	0.0433	0.0533	0.0400	0.1033	0.1400	0.0700	(0.5,0.1)
0.10	0.0933	0.0867	0.1000	0.0867	0.0900	0.1533	0.2133	0.1100	(0.5,0)
	0.0700	0.0833	0.0833	0.1000	0.1000	0.1700	0.2133	0.1133	(0.5,0.1)

P\\$\hat{\theta}$	0.8967	0.1500	2036	0.4045	0.3487	-0.0001	1.2203	1.1073	$(\rho,\gamma))$
	0.8954	0.1517	0.1962	0.4024	0.3468	1009	1.2178	1.1063	
0.01	0.0167	0.0133	0.0067	0.0100	0.0067	0.033	0.0467	0.0100	(0.9,0)
	0.0100	0.0067	0.0133	0.0267	0.0100	0.0400	0.0773	0.0167	(0.9,0.1)
0.05	0.0600	0.0633	0.0600	0.0567	0.0367	0.1033	0.1567	0.0600	(0.9,0)
	0.0400	0.0700	0.0400	0.0933	0.0733	0.0900	0.1533	0.0633	(0.9,0.1)
0.10	0.1067	0.0867	0.0900	0.0833	0.0933	0.1533	0.2000	0.1067	(0.9,0)
	0.1067	0.0900	0.0700	0.1633	0.1300	0.1333	0.2233	0.1133	(0.9,0.1)

Non-normality

Rrepetitions = 300, $\hat{\theta} = \frac{1}{300}\sum_{j=1}^{300}\theta_j^*$, $\sqrt{1.5} \simeq 1.2247$, $\sqrt{1.25} \simeq 1.1180$

The content of bracket is $(\rho_0\ ,\ \gamma_0)$

Table 3.8: $H_0 : \theta = \theta_0$, where ρ =0 ~ 0.9

$$\theta = (\rho, 0.2, 0.4, \gamma, \sqrt{1.5}, \sqrt{1.25})$$
$$\theta_0 = (\rho_0, 0.2, 0.4, \gamma_0, \sqrt{1.5}, \sqrt{1.25})$$

P\ $\rho_0 \overset{=}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	(ρ, γ)
0.01	0.0100	0.1433	0.6067	0.9333	1.0000	1.0000	1.0000	1.0000	(0,0)
	0.0133	0.1767	0.6467	0.9467	1.0000	1.0000	1.0000	1.0000	(0,0.1)
0.05	0.0600	0.2967	0.7833	0.9767	1.0000	1.0000	1.0000	1.0000	(0,0)
	0.0633	0.3400	0.8367	0.9867	1.0000	1.0000	1.0000	1.0000	(0,0.1)
0.10	0.1067	0.4733	0.8933	1.0000	1.0000	1.0000	1.0000	1.0000	(0,0)
	0.1167	0.3700	0.8400	0.9833	1.0000	1.0000	1.0000	1.0000	(0,0.1)
P\ $\rho_0 \overset{=}{\rightarrow}$	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	(ρ, γ)
0.01	0.9300	0.5200	0.0700	0.0133	0.1600	0.5400	0.9000	0.9833	(0.5,0)
	0.9900	0.7633	0.2567	0.0433	0.3633	0.8000	0.9867	0.9967	(0.5,0.1)
0.05	0.9867	0.7167	0.2200	0.0600	0.3100	0.7667	0.9667	0.9967	(0.5,0)
	0.9900	0.7633	0.2567	0.0433	0.3633	0.8000	0.9867	0.9967	(0.5,0.1)
0.10	0.9933	0.8133	0.3633	0.0933	0.4233	0.8400	0.9733	0.9967	(0.5,0)
	0.9967	0.8733	0.3800	0.0700	0.4500	0.8567	0.9900	1.0000	(0.5,0.1)
P\ $\rho_0 \overset{=}{\rightarrow}$	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	(ρ, γ)
0.01	1.0000	1.0000	1.0000	0.8800	0.1967	0.0167	0.3000	0.8933	(0.9,0)
	1.0000	1.0000	1.0000	0.9233	0.2567	0.0100	0.3867	0.9533	(0.9,0.1)
0.05	1.0000	1.0000	1.0000	0.9733	0.4167	0.0600	0.5367	0.9600	(0.9,0)
	1.0000	1.0000	1.0000	0.9900	0.4800	0.4400	0.6267	0.9967	(0.9,0.1)
0.10	1.0000	1.0000	1.0000	0.9900	0.5367	0.1067	0.6567	0.9867	(0.9,0)
	1.0000	1.0000	1.0000	0.9933	0.6067	0.1067	0.7033	0.9967	(0.9,0.1)

Non-normality

Rrepetitions = 300, $\hat{\theta} = \frac{1}{300}\sum_{j=1}^{300}\theta_j^*$, $\sqrt{1.5} \simeq 1.2247$, $\sqrt{1.2} \simeq 1.1180$

The content of bracket is $(\rho_0\ ,\ \gamma_0)$

Table 3.9: Case a: $lnwage_{it} = \rho\ lnwage_{i,t-1} + \ a_i + \gamma\ a_i \cdot lnwage_{i,t-1} + \varepsilon_{it}$

$$\begin{aligned} lnwage_{it} &= \rho\ lnwage_{i,t-1} + \ a_i + \gamma\ a_i \cdot lnwage_{i,t-1} + \varepsilon_{it}, \\ a_i &= \alpha_0 + \alpha_1\ lnwage_{i0} + c_i \end{aligned}$$

Coefficient	$\hat{\rho}$	$\hat{\alpha_0}$	$\hat{\alpha_1}$	$\hat{\gamma}$	$\hat{\sigma}_\varepsilon$	$\hat{\sigma}_a$
CMLE	-0.4951	0.9448	0.0758	0.8289	0.3493	0.0898
t-statistics	(-1.744)	(35.986)	(3.893)	(2.754)	(75.516)	(4.148)
log-likelihood value	302.614					

N=545, periods is 1980 $\sim$ 1987

Table 3.10: Case b:$lnwage_{it} = \vartheta\ lnwage_{i,t-1} + \ a_i + \gamma\ (a_i - \mu_a) \cdot lnwage_{i,t-1} + \varepsilon_{it}$

$$\begin{aligned} lnwage_{it} &= \vartheta\ lnwage_{i,t-1} + \ a_i + \gamma\ [a_i - \mu_a] \cdot lnwage_{i,t-1} + \varepsilon_{it}, \\ a_i &= \alpha_0 + \alpha_1\ lnwage_{i0} + c_i \end{aligned}$$

Coefficient	$\hat{\vartheta}$	$\hat{\alpha_0}$	$\hat{\alpha_1}$	$\hat{\gamma}$	$\hat{\sigma}_\varepsilon$	$\hat{\sigma}_a$
CMLE	0.3755	0.9448	0.0759	0.8285	0.3493	0.0898
t-statistics	(17.046)	(35.995)	(3.843)	(2.759)	(75.531)	(4.156)
log-likelihood value	302.614					

N=545, periods is 1980 $\sim$ 1987

Table 3.11: Case c:$lnwage_{it} = \rho\ lnwage_{i,t-1} + \delta_t\ d_t + \ a_i + \gamma\ a_i \cdot lnwage_{i,t-1} + \varepsilon_{it}$

$$\begin{aligned} lnwage_{it} &= \rho\ lnwage_{i,t-1} + \delta_t\ d_t + \ a_i + \gamma\ a_i \cdot lnwage_{i,t-1} + \varepsilon_{it}, \\ a_i &= \alpha_0 + \alpha_1\ lnwage_{i0} + c_i \end{aligned}$$

Coefficient	$\hat{\rho}$	$\hat{\alpha_0}$	$\hat{\alpha_1}$	$\hat{\gamma}$	$\hat{\delta_1}$	$\hat{\delta_2}$	$\hat{\delta_3}$
CMLE	0.0083	0.6565	0.1652	0.2567	0.2842	0.3093	0.3424
t-statistics	(0.0067)	(1.31)	(7.389)	(3.605)	(0.563)	(0.613)	(0.678)

Coefficient	$\hat{\delta_4}$	$\hat{\delta_5}$	$\hat{\delta_6}$	$\hat{\delta_7}$	$\hat{\sigma}_\varepsilon$	$\hat{\sigma}_a$
CMLE	0.4023	0.4328	0.4815	0.5356	0.3309	0.1841
t-statistics	(0.797)	(0.858)	(0.954)	(1.061)	(78.1)	(10.089)

N=545, periods is 1980 ~ 1987

Table 3.12: Case d:$lnwage_{it} = \vartheta\ lnwage_{i,t-1} + \delta_t\ d_t + \ a_i + \gamma\ (a_i - \mu_a) \cdot lnwage_{i,t-1} + \varepsilon_{it}$

$$\begin{aligned} lnwage_{it} &= \vartheta\ lnwage_{i,t-1} + \delta_t\ d_t + \ a_i + \gamma\ [a_i - \mu_a] \cdot lnwage_{i,t-1} + \varepsilon_{it}, \\ a_i &= \alpha_0 + \alpha_1\ lnwage_{i0} + c_i \end{aligned}$$

Coefficient	$\hat{\vartheta}$	$\hat{\alpha_0}$	$\hat{\alpha_1}$	$\hat{\gamma}$	$\hat{\delta_1}$	$\hat{\delta_2}$	$\hat{\delta_3}$
CMLE	0.21545	0.50248	0.223801	0.134065	0.599911	0.6541561	0.397765
t-statistics	(11.86)	(0.2724)	(8.9881)	(2.5400)	(0.3253)	(0.3547)	(0.2157)

Coefficient	$\hat{\delta_4}$	$\hat{\delta_5}$	$\hat{\delta_6}$	$\hat{\delta_7}$	$\hat{\sigma}_\varepsilon$	$\hat{\sigma}_a$
CMLE	0.422917	0.457085	0.517636	0.549881	0.330714	0.21901
t-statistics	(0.2293)	(0.2487)	(0.2807)	(0.2982)	(73.9500)	(12.0017)

N=545, periods is 1980 ~ 1987

Table 3.13: Case e: $lnwage_{it} = \rho\ lnwage_{i,t-1} + \beta\ union_{it} +\ a_i + \gamma\ a_i \cdot lnwage_{i,t-1} + \varepsilon_{it}$

$$\begin{aligned} lnwage_{it} &= \rho\ lnwage_{i,t-1} + \beta\ union_{it} +\ a_i + \gamma\ a_i \cdot lnwage_{i,t-1} + \varepsilon_{it}, \\ a_i &= \alpha_0 + \alpha_1\ lnwage_{i0} + \alpha_2\ \overline{union_i} + c_i \end{aligned}$$

Coefficient	$\hat{\rho}$	$\hat{\beta}$	$\hat{\alpha_0}$	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\gamma}$	$\hat{\sigma}_\varepsilon$	$\hat{\sigma}_a$
CMLE	-0.4771	0.0507	0.9430	0.0820	0.0244	0.7757	0.3518	0.0987
t-statistics	(-1.861)	(2.049)	(33.093)	(3.923)	(1.248)	(2.819)	(67.366)	(4.279)
log-likelihood value	85.6199							

N=545, periods is 1980 ~ 1987

Table 3.14: Case f: $lnwage_{it} = \vartheta\ lnwage_{i,t-1} + \beta\ union_{it} + a_i + \gamma[a_i - \mu_a] \cdot lnwage_{i,t-1} + \varepsilon_{it}$

$$\begin{aligned} lnwage_{it} &= \vartheta\ lnwage_{i,t-1} + \beta\ union_{it} +\ a_i + \gamma\ [a_i - \mu_a] \cdot lnwage_{i,t-1} + \varepsilon_{it}, \\ a_i &= \alpha_0 + \alpha_1\ lnwage_{i0} + \alpha_2\ \overline{union_i} + c_i \end{aligned}$$

Coefficient	$\hat{\vartheta}$	$\hat{\beta}$	$\hat{\alpha_0}$	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\gamma}$	$\hat{\sigma}_\varepsilon$	$\hat{\sigma}_a$
CMLE	0.3480	0.0511	0.9425	0.0820	0.0240	0.7773	0.3518	0.0985
t-statistics	(13.726)	(2.065)	(33.091)	(3.909)	(1.235)	(2.814)	(67.343)	(4.265)
log-likelihood value	85.6199							

N=545, periods is 1980 ~ 1987

CHAPTER 4

CMLE For Logit Model With Individual Heterogeneity

4.1 Introduction

In the previous chapters, we considered estimation of the AR(1) panel data model with unobserved heterogeneity. In the standard model with additive heterogeneity, transformations exist that can be combined with instrumental variables estimation to produce consistent estimators. (As we discussed in Chapter 1, the usual within estimator is not consistent with fixed T.) Even in this simple model , however, the conditional maximum likelihood estimator has some advantages. For one, it is generally more efficient than method of moments estimators that do not make assumptions on the distribution of the initial condition.

In Chapter 3 we considered estimation of the AR(1) model when the unobserved heterogeneity and the lagged dependent variable possibly interact. As shown there, the usual IV estimators that are consistent in the model with only additive heterogeneity are no longer consistent when the autoregressive coefficient depends on the unobserved heterogeneity. Nevertheless, the conditional maximum likelihood approach does produce consistent estimators (under a normality assumption and, perhaps, more generally). In the empirical application to a dynamic wage equation, the interaction between lagged log wage and the unobserved heterogeneity was statistically and practically important.

In this chapter we turn to a model where conditional maximum likelihood methods are indispensable: the dynamic logit model with unobserved heterogeneity. The fact that the logit model is nonlinear makes dealing with a lagged dependent variable even much more difficult than in Chapter 2 and 3. First, with small T, we cannot simply treat the unobserved effects as parameters to estimate. Even without a lagged dependent variable, the incidental parameter problem (Neyman and Scott [1948]) caused inconsistent estimation of the parameters. Secondly, as with the linear AR(1) model, the inclusion of lagged dependent variable is very difficult to characterize theoretically, but the intuition is the same as for the linear model. Plus, for large N, treating the unobserved heterogeneity as parameters to estimate is computationally burdensome.

In this chapter I show how to implement conditional maximum likelihood estimation, following the general treatment of Wooldridge (2000). As in the previous chapters, this entails modeling the distribution of the unobserved heterogeneity conditional on the initial condition and any strictly exogenous variables. Nevertheless, we need not find a steady state distribution for initial condition, and we need not approximate this distribution. An added benefit is that we can consistently estimate the average partial effects - that is, the partial effect averaged across the distribution of the heterogeneity - rather than just the parameter. Thus, while this approach is almost fully parametric, it delivers estimates of interesting quantities that semiparametric approaches cannot.

The plan of this chapter is as follows. Section 2 applies the CMLE to a basic dynamic logit model with unobserved effects. In this section we construct the conditional likelihood function to obtain the conditional maximum likelihood estimators. Section 3 we set up Monte Carlo studies to examine the performance of conditional maximum likelihood estimator as $N \to \infty$ with small T. Section 4 investigates empirical example of union membership and calculate the average partial effects. Finally, we make some concluding remarks.

4.2 The CMLE for Dynamic Logit Model with Unobserved Heterogeneity

4.2.1 Estimation of Fixed Effects Model

I consider the dynamic logit model with unobserved heterogeneity as follows:

$$
\begin{aligned}
P(y_{it} &= 1 | y_{i0}, \ldots, y_{i,t-1}, x_i, a_i) \\
&\equiv F(y_{it}) \\
&\equiv \Lambda(\rho y_{i,t-1} + x_{it}\beta + a_i) \\
&\equiv \frac{\exp(\rho y_{i,t-1} + x_{it}\beta + a_i)}{1 + \exp(\rho y_{i,t-1} + x_{it}\beta + a_i)} \qquad t = 1, \ldots, T; T > 2, i = 1, \ldots, N,
\end{aligned}
\tag{4.1}
$$

where a_i is an individual-specific effect that may depend on the exogenous explanatory variables $x_i \equiv (x_{i1}, \ldots, x_{it})$ in an arbitrary way and where y_{i0} is the initial value of the response variable. If we assumed $F(y_{it})$ is a linear function we would have a dynamic linear probability model and we would apply IV methods discussed in Chapter 1 to estimate the parameters. The LPM, however, has inherent defects because the response probability might not be constrained between 0 and 1, and the true response effect is probably not constant. Theoretically, the inability to obtain the consistent estimator of a_i will not rule out the possibility of obtaining a consistent estimator of β if equation (4.1) is in the form of static linear-regression model, $y_{it} = x_{it}\beta + a_i + \varepsilon_{it}$, because the estimation of β and a_i are asymptotically independent (Hsiao [1986]). Even more, we can obtain an appropriate transformation to remove the effect of a_i in regressive model provided that we properly interpret the initial conditions. Unfortunately, the same things can not be said for the non-linear case because of the estimation of β or ρ and a_i are not independent of each other. The inconsistency of a_i is transmitted into the estimator β or ρ. For example, as discussed in Chapter 3, when the lagged dependent variable interacts with the unobserved heterogeneity, no transformations immediately suggest themselves to eliminate the effect of unobserved heterogeneity.

Just as in the dynamic linear-regression model, the problem of initial conditions for the dynamic logit model with unobserved heterogeneity must be resolved before we can consistently estimate the parameters generating the stochastic process. The effect is very difficult to characterize theoretically, but the intuition is the same as for the linear model. Even for a static logit model, ($\rho = 0$ in (4.1), $P(y_{it} = 1|x_i, a_i) = \Lambda(x_{it}\beta + a_i)$, the assumption of non-random unobserved effect, a_i, means we need to estimate both β and a_i which are unknown parameters. When T tends to infinity, the MLE is consistent. However, as we know, T is usually small for panel data, in which case we have an incidental parameters.

Let us illustrate the inconsistency of the MLE for β in the static logit model in the following (Hsiao [1992]). The log-likelihood function for the static model (4.1) given $\rho = 0$ is

$$\log L = -\sum_i \sum_t \log[1 + \exp(x_{it}\beta + a_i)] + \sum_i \sum_t y_{it}(x_{it}\beta + a_i). \qquad (4.2)$$

For simplicity, we assume $T = 2$, one explanatory variable, with x_{i1} =0 and $x_{i2} = 1$. Then the first-derivative equations are

$$\frac{\partial \log L}{\partial \beta} = \sum_i \left[-\frac{\exp(\beta + a_i)}{(1 + \exp(\beta + a_i)} + y_{i2} \right] = 0, \qquad (4.3)$$

$$\frac{\partial \log L}{\partial a_i} = \sum_{t=1}^{2} \left[-\frac{\exp(\beta x_{it} + a_i)}{1 + \exp(\beta x_{it} + a_i)} + y_{it} \right] = 0, \qquad (4.4)$$

Solving (4.4), we have

$$\begin{aligned} a_i &= \infty \quad \text{if } y_{i1} + y_{i2} = 2, \\ a_i &= -\infty \quad \text{if } y_{i1} + y_{i2} = 0, \\ a_i &= -\tfrac{\beta}{2} \quad \text{if } y_{i1} + y_{i2} = 1. \end{aligned} \qquad (4.5)$$

Inserting (4.5) into (4.3) and letting n_1 denote the number of individuals with $y_{i1} + y_{i2} = 1$ and n_2 denote the number of individuals with $y_{i1} + y_{i2} = 2$, we have

$$\sum_i^N \frac{\exp(\beta + a_i)}{(1 + \exp(\beta + a_i)} + y_{i2} = n_1 \frac{\exp(\beta/2)}{1 + \exp(\beta/2)} + n_2 = \sum_i^N y_{i2}. \qquad (4.6)$$

Therefore,

$$\hat{\beta} = 2\{\log(\sum_i y_{i2} - n_2) - \log(n_1 + n_2 -)\}.$$

By a law of large number

$$\operatorname*{plim}_{N\to\infty} \hat{\beta} = 2\beta,$$

which is not consistent because

$$\operatorname*{plim}_{N\to\infty} \tfrac{1}{N}(\sum_{i=1}^N y_{i2} - n_2) = \tfrac{1}{N}\sum_{i=1}^N \frac{\exp(\beta + a_i)}{(1 + \exp(a_i))(1 + \exp(\beta + a_i))},$$

$$\operatorname*{plim}_{N\to\infty} \tfrac{1}{N}(n_1 + n_2 - \sum_{i=1}^N y_{i2}) = \tfrac{1}{N}\sum_{i=1}^N \frac{\exp(\beta + a_i)}{(1 + \exp(a_i))(1 + \exp(\beta + a_i))},$$

Thus we have incidental-parameter problem in that there is only a limited number of observations to estimate a_i (Neyman and Scott [1948]). It is meaningless that any estimation of a_i if we intend to judge the estimators by the large-sample properties ($N \to \infty$).

The nonlinear panel data model with individual heterogeneity may be estimated by the semiparametric approach which allows us to make use of the linear structure of the latent variable equations such that the individual-specific unobserved effect can be eliminated by the differencing transformation and the like and hence the lack of knowledge of a_i no longer affects the estimation of parameters of interest (Manski [1987]). Recently, Honoré and Kyriazidou (2000) derive effective moment conditions for the unobserved effects logit model with one-period lagged dependent variable from an objective function that identify the parameters. The interesting advantages of semiparametric approaches is to allow estimation of parameters without specifying distributions for the unobserved effects, although the estimators may not possibly converge at the rate $\sqrt{N}$ (Hahn [1997]). The nature of semiparametric approaches, nevertheless, can not suggest the estimators of partial effects on mean responses.

The nonlinear model with unobserved effects might be estimated by a random effect approach. Such an approach requires the specification of the statistical relationship between the observed covariates and unobserved permanent individual heterogeneity. Furthermore, it entail specify the distribution of initial condition if the list of explanatory variables include the lagged dependent variables. The inherent defect is the misspecification of these distributions. We are to use the parametric to solve the dynamic logit model with unobserved effects by specifying the conditional distribution for the unobserved heterogeneity and hence we also incur the question, which misspecification of this distribution generally leads to inconsistent parameter estimates. We, nevertheless, have set up a simple conditional maximum likelihood estimators and moreover, the quantities of interest in nonlinear case can be obtained based on the assumptions employed, in particular, the partial effects on the mean response, averaged across the population distribution of the unobserved heterogeneity.

4.2.2 Conditional Maximum Likelihood Estimator

First, we construct the conditional likelihood function for the conditional maximum likelihood estimator in dynamic logit model with unobserved effects (4.1), given $\beta = 0$. Generally, we assume that ε_{it} is symmetrically distributed about zero, which means that 1- $G(-z) = G(z)$ for all real numbers z. We make the assumptions as follows:

Assumption 4.1 ε_{it} is independent of $\varepsilon_{i,t-1}, \ldots, \varepsilon_{i1}, y_{i0}$, and a_i.

Assumption 4.2 $a_i|y_{i0} \sim \text{Normal}(\alpha_0 + \alpha_1\ \mathrm{y}_{i0}, \sigma^2_{\mathrm{a}})$.

According to Assumption 4.1, the conditional density function is as follows:

$$f(y_{it}|y_{i,t-1}, \ldots, y_{i0}, a_i) = \Lambda(\rho\ y_{i,t-1} +\ a_i)^{y_{it}}\ (1 - \Lambda(\rho\ y_{i,t-1} +\ a_i))^{(1-y_{it})}, \tag{4.7}$$

where

$$\Lambda(\rho\ y_{i,t-1} +\ a_i) = \frac{\exp(\rho\ y_{i,t-1} +\ a_i)}{1 + \exp(\rho\ y_{i,t-1}\ + a_i)}$$

and hence the density function of T-period observations of cross-section i:

$$f(y_{iT}, \ldots, y_{i1}|y_{i0}, a_i) = \prod_{t=1}^{T} \Lambda(\rho\ y_{i,t-1} +\ a_i)^{y_{it}}\ (1 - \Lambda(\rho\ y_{i,t-1} +\ a_i))^{(1-y_{it})} \tag{4.8}$$

To obtain a conditional log-likelihood function for the T-period observations of cross-section i, we specify a distribution for the unobserved effects $h(a_{i0}|y_{i0}; \alpha)$ and then we obtain the likelihood function of T-periods of cross-section i conditioning on y_{i0} by equation (4.8) and $h(a_{i0}|y_{i0}; \alpha)$ as follows:

$$\ell\ (y_{iT}, \ldots, y_{i0}; \theta)\ = \\ \log \int_{-\infty}^{\infty} \left[\prod_{t=1}^{T} \Lambda(\rho\ y_{i,t-1} +\ a)^{y_{it}}\ (1 - \Lambda(\rho\ y_{i,t-1} +\ a))^{(1-y_{it})}\right] h(a_i|y_{i0}; \alpha)\ \mathrm{d}\,a, \tag{4.9}$$

where $\theta = (\rho\ ,\ \alpha)$. We can solve the conditional maximum likelihood estimator by maximizing the sum of equation (4.9) across i=1,..., N with respective to θ. We write the maximizing problem as follows:

$$\max_{\theta} \sum_{i=1}^{N} \ell\ (y_{iT}, \ldots, y_{i0}; \theta) \tag{4.10}$$

In the chapter 2, we have discussed the consistency of conditional maximum likelihood function. Equation (4.9) satisfies the generic form (2.9) and thus satisfies the inequality equation (2.15), Kullback-Leibler information inequality and equation (2.16), this ensures that the true parameters θ_0 solve the relevant population maximization problem, but they still might not be the unique solutions. For identification, we must assume that the inequality is strict. According to Assumption 4.1 and 4.2, equation (4.10) can be rewritten as follows:

$$\log \int_{-\infty}^{\infty} \left[\prod_{t=1}^{T} \Lambda(\rho\ y_{i,t-1} +\ a)^{y_{it}}\ (1 - \Lambda(\rho\ y_{i,t-1} +\ a))^{(1\ -\ y_{it})}\right] \cdot \\ (\frac{1}{\sqrt{2\pi\sigma_a^2}}) \exp(\frac{-1}{2}(\frac{c}{\sigma_a})^2)\, \mathrm{d}\, c, \tag{4.11}$$

where $a_i\ =\ \alpha_0 + \alpha_1 y_{i0} + c_i$ and $c_i\ \sim\ \text{Normal}(0, \sigma_{\text{a}}^2)$ from Assumption 4.2.

It is impossible to reach a formula of closed form for the estimator by directly solving out the first conditions of sum of equation (4.10) across i from 1 to N. We need to employ numerical methods. Under Assumption 4.1 and 4.2, the formula for the evaluation of the necessary integral of (4.11) is the Hermite integral formula $\int_{-\infty}^{\infty} \mathrm{e}^{-z^2} g(z)\, \mathrm{d}\, z\ =\ \sum_{j=1}^{K} w_j g(z_j)$, where K is the number of evaluation points, w_j is the weight given to the jth evaluation point, and $g(z_j)$ is $g(z)$ evaluated at the jth point of z (Butler and Moffitt [1982]). This formula is appropriate to our problem because the normal density h in equation (4.11) contains a term can be expressed as a form of e^{-z^2} and the function of $g(z)$ is, in our case, the density function of T-period observations of cross-section i. Without finding a specific distribution for

$h(a_i|y_{i0})$, an integration by Gaussian method might be needed. For simplification of numerical calculation, we always specify a distribution for $h(a_i|y_{i0})$ to fit for the Hermite integral formula in the simulation.

The previous analysis is limited on the time-series observation of cross section, $\{y_{it}\}_{i=1,\dots,N}^{t=1,\dots,T}$, without the exogenous variables. In most of application, we add other exogenous variables to study the response of the (explanatory variables to the future and further the feedback from the unexpected movements in the outcome variable to future values of the explanatory variables). Some experimental case where the variable will be in the control of a researcher. Honoré and Kyriazidou (2000) give a restriction on x_{it} to identify the parameters. For example, if x_{it} represents some program participation, $x_{it} = x_{i,t-1}$ means that the status of participation will not change for successive periods.

We consider the model for union membership with unobserved heterogeneity. The key explanatory variables, such as the school year or education diploma, graduate or non-graduate, is more or less related to the union membership. A variable such as a person's age can be thought of as strictly exogenous variable if we just study the male youth. Using the general framework of CMLE in Chapter 2 under the strict exogeneity, we specify distributions for $(y_T, \dots, y_1)$ given $(x_T, \dots, x_1,\ \ y_0)$and a_i given $(x_T, \dots, x_1,\ \ y_0)$ as $\mathrm{D}(y_T, \dots, y_1|x_T, \dots, x_1, y_0)$ and $H(a|x_T, \dots, x_1, y_0)$ in respective. The parameterized density function for the distributions are $f(y_T, \dots, y_1|x_T, \dots, x_1, y_0, \delta)$ and $h(a_i|x_T, \dots, x_1, y_0, \delta)$ in respective. In practice, $E(a_i|x_{iT}, \dots, x_{i1}, y_{i0})$ is assumed to be in a function of $(x_{iT}, \dots, x_{i1}, y_{i0})$. We assume that $E(a_i|x_{iT}, \dots, x_{i1}, y_{i0}) = \alpha_0 + \alpha_1 y_{i0} + \alpha_2 \overline{x}_i$, where $\overline{x}_i$ is a linear combination of $(x_{i,T}, \dots, x_{i1})$, $\overline{x}_i = \sum_{t=1}^{T} \tau_t x_{it}$. We assume that τ_t is equal to 1 for $t = 1, \dots, T$ to decrease the number of identification for parameters. We make assumptions about

ε_{it} and a_i as follows:

Assumption 4.3 $\varepsilon_{it}|y_{i,t-1},\ldots,y_{i0},x_{iT},\ldots,x_{i1},a_i \sim Logit(0,\frac{\pi^2}{3})$.

Assumption 4.4 $a_i|x_{iT},\ldots,x_{i1},y_{i0} \sim \text{Normal}\ (\alpha_0+\alpha_1\ \mathrm{y_{i0}}+\bar{\mathrm{x}}_\mathrm{i}\ \alpha_2,\sigma_\mathrm{a}^2)$.

According to Assumption 4.3 and 4.4, $\mathrm{f}(y_T,\ldots,y_1|x_T,\ldots,x_1,y_0,\delta)$ is equal to $\prod_{t=1}^T \Lambda(\rho\ y_{i,t-1}+\ x_{it}\beta+\ a_i)^{y_{it}}\ (1-\Lambda(\rho\ y_{i,t-1}+\ x_{it}\beta+\ a_i)^{(1\ -\ y_{it})}$. Equation (4.9) can be rewritten as follows:

$$\ell\ (y_{iT},\ldots,y_{i0},x_{iT},\ldots,x_{i1};\theta)\ =$$
$$\log\int_{-\infty}^{\infty}\left[\prod_{t=1}^T\Lambda(\rho\ y_{i,t-1}+\ x_{it}\beta+\ a)^{y_{it}}\ (1-\Lambda(\rho\ y_{i,t-1}+\ x_{it}\beta+\ a)^{(1\ -\ y_{it})}\right]$$
$$h(a|x_{i1},\ldots,x_{iT},y_{i0};\alpha)\ \mathrm{d}\,a \tag{4.12}$$

Replacing the above equation into the equation (4.12) and then solve out the maximization of the objection function (4.12).

If X_i is not strictly exogenous, we can apply the suggestion of Wooldridge (2000a) as follows. To parameterize $g(x_t|X_{t-1},z_t,a;\lambda_0)$, where z_t is strictly exogenous and build up the joint density of (Y_t,X_t) given (Z_T,Y_{t-1},X_{t-1},a) and then apply the same procedure as discussed previously to set up a log-likelihood function. We can use the numerical method to solve out the CMLE.

4.3 Simulation Evidence

In order to investigate the performance of maximum-likelihood estimators given the initial value, we conducted Monte Carlo studies. We divide this section by two subsection: one is for the model without exogenous variable; the other is for the model with strictly exogenous variable. We use the MLE software of Gauss to do our simulation for the conditional maximum likelihood estimator. As the discussion in previous chapter, the feasible computation of the Hermite integral depends on the number of evaluation points at which the integrand must be evaluated for accurate approximation. Several evaluations of the integral using seven periods of arbitrary values of data and coefficients on two right-hand-side variables shows that the value of K is chosen to be 21 is highly accurate. Although the value of K determines the accuracy of the calculation of integral, we don't discuss the relation of K and the evaluation of integral as Butter and Moffitt (1982) did. K=21 is highly accurate for the evaluation of integral (4.12). We repeat the maximization of the model of interest in the following for 500 hundreds. The notations for the simulation are as follows:

1. θ^* means the conditional maximum likelihood estimators in each iteration.

2. $\hat{\theta} = \frac{1}{500}\sum_{j=1}^{500}\theta_j^*$.

3. θ means true value of parameter, where θ=$(\rho, \alpha_0, \alpha_1, \sigma_a)$= $(\rho, 0.2, 0.4, \sqrt{1.2})$ in the model without exogenous variables or θ=$(\rho, \beta, \alpha_0, \alpha_1, \alpha_2, \sigma_a)$= $(\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{1.2})$ in the model with strictly exogenous variables.

4.3.1 The Model Without Exogenous variables

Let the true value of ρ be 0, 0.25, 0.5, 0.75, 0.9, and 0.95. We calculate the frequency of rejecting the hypothesis of $H_0 : \theta = \theta_0$ to examine the performance of the conditional maximum likelihood estimator. With the same procedure, we focus on the estimator,$\hat{\rho}$ by calculating the frequency of rejecting the hypothesis of $H_0 : \rho = \rho_0$ under different true value of ρ, where ρ_0 is 0, 0.25, 0.5, 0.75, 0.9, and 0.95. The results are reported in Table 4.1 - 4.4. We begin with the true model as follows:

$$\begin{aligned} y_{it}^* &= \rho_0\ y_{i,t-1} + \ a_i + \ \varepsilon_{it}, \\ y_{it} &= 1[y_{it}^* > 0], \end{aligned} \tag{4.13}$$

$$i = 1, \ldots, N, \quad t = 1, \ldots, T,$$

where $a_i = 0.2 + 0.4\ y_{i0} + c_i$. The c_i comes from Normal(0, 1.2). According to Assumption 4.1, the inverse function of the logistic function, $u = \frac{e^{\varepsilon}}{1 - e^{\varepsilon}}$ is $\ln(\frac{u}{1+u})$. Therefore, we generate the $\varepsilon_{it} = \log(\frac{u_{it}}{1+u_{it}})$, where u_{it} comes from the uniform distribution of [0,1]. The conditional likelihood function (4.11) can be rearranged as follows:

$$\log \int_{-\infty}^{\infty} \left[\prod_{t=1}^{T} \frac{\exp[y_{it}(\rho\ y_{i,t-1} + a)]}{1 + \exp(\rho\ y_{i,t-1} + a)} \right] (\frac{1}{\sqrt{2\pi\sigma_a^2}}) e^{\frac{-1}{2}(\frac{a-\mu_a}{\sigma_a})^2} \mathrm{d}\, a, \tag{4.14}$$

Let z_i to be $\frac{a_i - \mu_a}{\sqrt{2}\sigma_a}$ and replace a_i with $\alpha_0 + \alpha_1\ y_{i0} + \sqrt{2}\sigma_a z_i$ into the function (4.14). Therefore the conditional likelihood function of cross section i can be re-written as follows:

$$\log \int_{-\infty}^{\infty} \left[\sqrt{\pi}^{-T} \prod_{t=1}^{T} \frac{\exp[y_{it}(\rho\ y_{i,t-1} + \alpha_0 + \alpha_1\ y_{i0} + \sqrt{2}\sigma_a z)]}{1 + \exp(\rho\ y_{i,t-1} + \alpha_0 + \alpha_1\ y_{i0} + \sqrt{2}\sigma_a z)} \right] e^{-z^2} \mathrm{d}\, z. \tag{4.15}$$

The integral of function (4.15) can be approximated by the Hermite integral formula: $\int_{-\infty}^{\infty} g(z) e^{-z^2} \mathrm{d}\, z = \sum_{j=1}^{K} w_j g(z_j)$. The likelihood function of (4.15) can be expressed in the form of Hermite integral formula as follows:

$$\log \int_{-\infty}^{\infty} g(z) e^{-z^2} \mathrm{d}\, z \simeq \log \sum_{j=1}^{K} w_j g(z_j), \tag{4.16}$$

where

$$g(z_i) = \sqrt{\pi}^{-T} \prod_{t=1}^{T} \frac{\exp\left[y_{it}(\rho\ y_{i,t-1} + \alpha_0 + \alpha_1\ y_{i0} + \sqrt{2}\sigma_a z_i)\right]}{1 + \exp(\rho\ y_{i,t-1} + \alpha_0 + \alpha_1\ y_{i0} + \sqrt{2}\sigma_a z_i)}.$$

We maximize the sum of the likelihood function (4.16) away from the constant term across i from 1 to N to obtain the estimators as follows:

$$\max \sum_{i=1}^{N} \mathit{log} \left[\sum_{j=1}^{K} w_j \left(\prod_{t=1}^{T} \frac{\exp\left[y_{it}(\rho\ y_{i,t-1} + \alpha_0 + \alpha_1\ y_{i0} + \sqrt{2}\sigma_a z_{ij})\right]}{1 + \exp(\rho\ y_{i,t-1} + \alpha_0 + \alpha_1\ y_{i0} + \sqrt{2}\sigma_a z_{ij})} \right) \right]. \tag{4.17}$$

In the simulation, we set the number of evaluation point, K to be 21. We examine the assumption about the conditional a_i: normality and non-normality. The result of Table 4.1 and 4.5 is under the normality assumption of a_i, while Table 4.3 and 4.7 is under the non-normality. We assume the t-distribution with freedom 10 to explain the non-normality assumption on a_i. Table 4.1 reports the CMLE estimates for the data generated from the true model. We repeat 500 times for the same procedure of maximizing the objective function (4.17) to obtain the CMLE estimates.

To examine the power test for the CMLE estimators, with 500 repetitions, we calculate how many times the hypothesis of $H_0 : \theta = \theta_0$ will be rejected under a certain power value, 0.1, 0.05, 0.10 respectively. For example, in the second column of Table 4.1, the number of bracket is the average value of 500 estimates, ρ^*; the values of the second to fourth row represent the p-value 0.004, 0.046, and 0.09 under the power 0.01, 0.05, 0.1 in respective when true value of ρ is zero. According to the result of the simulation, the CMLE estimators perform well. Similar to the linear case in Chapter 2, the value of ρ is likely to be rejected when the true value of ρ is getting further away from zero.

The response probability of interest is mainly related to the ρ, so we construct another Table 4.2- (i) - (vi) to examine the hypothesis $H_0 : \rho = \rho_0$, where ρ_0 is 0, 0.25, 0.5, 0.75, 0.9, and 0.95 under the different true value of ρ, 0, 0.25, 0.5, 0.75, 0.9, 0.95. For example, when the true value of $\rho = 0.25$ and ρ_0=0.6, the p-value is 0.2620 in Table 4.2- (iii) under the level of p-value, 0.01. Table 4.2- (vi), the true value of $\rho = 0.9$ and the $\rho = 0.55$, the p-value is 0.24 under the same level of p-value. This numerical evidence shows that the estimates away from the true value is more likely to be rejected when the true value of ρ is getting closer to zero. When we decrease the tolerance of confidence to 0.05 or 0.1, there is no crucial difference of p-value whatever the true value of ρ is.

4.3.2 The Model With Exogenous Variables

We put an end to the simulation with including an exogenous variable x_{it}. We still hold the strict exogeneity assumption. From the likelihood function (4.12), we construct the log liklihood function as follows:

$$\log \int_{-\infty}^{\infty} \left[\sqrt{\pi}^{-T} \prod_{t=1}^{T} \frac{\exp[y_{it}(\rho\, y_{i,t-1} + \ x_{it}\beta + \mu_a + \sqrt{2}\sigma_a z)]}{1 - \exp(\rho\, y_{i,t-1} + \ x_{it}\beta + \mu_a + \sqrt{2}\sigma_a z)} \right] \mathrm{e}^{-z^2}\, \mathrm{d}\, z. \qquad (4.18)$$

where $\mu_a = \alpha_0 + \alpha_1\ y_{i0} + \ \overline{x}_i\alpha_2$. We maximize the sum of the objective function (4.18) away from the constant term across i=1..., N as follows:

$$\max \sum_{i=1}^{N} \log \left[\sum_{j=1}^{K} w_j \left(\prod_{t=1}^{T} \frac{\exp\left[y_{it}(\rho\, y_{i,t-1} + \ x_{it}\beta + \mu_a + \sqrt{2}\sigma_a z_{ij})\right]}{1 - \exp(\rho\, y_{i,t-1} + \ x_{it}\beta + \mu_a + \sqrt{2}\sigma_a z_{ij})} \right) \right]. \qquad (4.19)$$

According to Assumption 4.3 and 4.4, we generate the ε_{it} and c_i as the former model do. We report the results in Table 4.5 - 4.8. Because we assume that x_{it} is strictly exogenous, it doesn't matter form which logic distribution x_{it} generates. We assume the x_{it} is continuous variable coming from the standard normal distribution. Table 4.6 shows that except the $\hat{\beta}$ the other estimates have the similar property of the former model. In Table 4.6, $\hat{\beta}$ is not significantly different from zero. Under the power 0.01,0.05 and 0.10, the hypothesis of $H_0 : \beta = 0$ can not be significantly rejected when the true value of β is 0.15. We calculated its relevant p-values of the test are 0.006, 0.036 and 0.09 under the power 0.01, 0.05, and 0.1 in respective. It might be the fact that we assume the conditional mean of unobserved effects $a_i = \alpha_0 + \alpha_1\ y_{i0} + \alpha_2\ \overline{x}_i$, where the $\overline{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$. The α_2 dominates the effect of x_{it} and accounts for most of its effect. This can be explained by the fact that the $hat\alpha_2$ is 0.523 while the true value of α_2 is 0.35. We calculate the p-values of $H_0 : \alpha_2 = 0$ are 0.3640, 0.6180 and 0.7640 under the power 0.01, 0.05, 0.1, in respective. That is we must pay much attention on the specification of $h(a_i|x_{iT}, \dots, x_{i1}, y_{i0})$.

4.4 Empirical Example

Statistical models developed for analyzing cross-sectional data essentially ignore individual differences and treat the aggregate of the individual effect and the omitted-variable effect as an incidental event. In this section we use the data from Vella and Verbeek (1998) to study the status of labor union membership. Such a Panel data make it possible, through the knowledge of the intertemporal dynamics of a worker who joins the labor union, to separate a model of individual behavior from a model of average behavior of a group of individuals. In particular, we might assume that the heterogeneity across cross-sectional units is time-invariant, and these individual-specific effects are captured by decomposing the error as $a_i + \varepsilon_{it}$. We always treat a_i as random to prevent the problem of incidental parameters. And the application of conditional maximum likelihood estimation into the model make it possible to do without restrictions on $union_{i0}$.

The existence of such unobserved time-invariant components allows individuals who are homogenous in terms of their observed characteristics to be heterogenous in response probabilities, $F(y_{it})$. For example, heterogeneity implies that the sequential-participation behavior of a worker, $F(union_{it})$, within a group of observationally homogenous worker differs systematically from the average behavior of the group, $\int F(union_{it}) \mathrm{d} H(a|union_{i0})$, $H(a|union_{i0})$ gives the population probability for a conditional the initial status of labor-union membership.

We use the data from Vella and Verbeek (1998) to study the conditional maximum likelihood estimator in estimating dynamic logit model using observations draw from a time series of cross sections. These data are for young males taken from the National Longitudinal Survey (Youth Sample) for the period 1980-87. We estimate a dynamic model for labor union status. Each of the 545 men in the sample worked in every year from 1980 through 1987. When a worker is a member of a labor union we set the status variable is one; when a worker is not a member of a labor union the status variable is zero. We examine the response probability of dependent variable, union membership over time series of cross section. For example, how do union membership of the past affect the probability of keeping the labor union membership at present, the amount of state dependence. We express the corresponding latent variable model as that $union^*_{it} = \rho\ union_{i,t-1} + a_i + c_i$ and we set up a logit model under Assumption 4.1 and 4.2 as follows:

$$\begin{aligned} P(union_{it} &= 1|union_{i,t-1}, \dots, union_{i0}, a_i) \\ &= P(\varepsilon_{it} > -(\rho\ union_{i,t-1} + a_i)|union_{i,t-1}, \dots, union_{i0}, a_i) \\ &= 1 - \Lambda(-(\rho\ union_{i,t-1} + a_i)) \\ &= \Lambda(\rho\ union_{i,t-1} + a_i), \end{aligned} \tag{4.20}$$

where $union_{it} = \begin{cases} 1 & , union^*_{it} > 0 \\ 0 & , otherwise. \end{cases}$

From equation (4.20), the partial effect of $union_{i,t-1}$ on the response probability is $(\Lambda(\rho + a_i) - \Lambda(a_i))$. Therefore, the conditional partial effect of $union_{i,t-1}$ on the response probability depends on the $union_{i,t-1}$ through the quantity $g(\rho\ union_{i,t-1} + a_i)$, meaning the difference of $\Lambda(\rho\ union_{i,t-1} + a_i)$ with respective to $union_{i,t-1}$, $\Lambda(\rho + a_i) - \Lambda(a_i)$. In Table 4.9, through the CMLE the estimates of ρ, α_0, and α_1 are 1.4923, -3.2775, and 2.669. These estimates are significantly different from zero. The estimated mean of unobserved heterogeneity a_i *is* $(-3.2775 + 2.669 \cdot union_{i0})$.

$\hat{a}_i = \hat{E}(a_i | union_{i0})$ is greater than -3.2775 and less than -0.6085.

That is , empirically, the individual unobserved heterogeneity of a worker tends to decrease the probability of keeping union membership, so the quantifying unobserved heterogeneity have negative effect on the response probability of union membership when we quantify the unobserved heterogeneity. The significant amount of unobserved effects means that previous membership of labor union appears to be a determinant of future membership mostly because it is a proxy for temporally persistent unobservables that determines the choice. If there is no other exogenous explanatory variable, the result shows that the effect of temporally persistent unobservables that determines joining the labor union or not is significant.

In other words, a worker participate in the union not just because he used to be a membership of the union; on the contrary, he might join the union in accordance with his own preference or some thing like the unobserved individual specific persistent heterogeneity. The result is consistent to the empirical result of Chapter 2, which the union membership accounts for not much of the wage rate. The estimated response of the current union membership into the probability of keeping union membership in the future is measure by $(\Lambda(\hat{\rho} - 0.6085) - \Lambda(-0.6085))$ instead of $\hat{\rho}$ when $union_{i0} = 1$. The estimated state dependence for a person with average of a_i is measured by the value of ($\Lambda(1.49 + \hat{\mu}_a)$ - $\Lambda(\hat{\mu}_a)$), equal to 0.1793, where $E(\hat{a_i}) = \alpha_0 + \alpha_1 y_0$, and $y_0 = \frac{1}{545}\sum_{i=0}^{545} y_{i0}$. Replacing $\hat{\mu}_a$ with the lower and upper bounds: $\underline{\hat{\mu}_a}$, -3.278 and $\overline{\hat{\mu}_a}$, -0.609, the range of state dependence effects is the interval of [-0.6665,-0.1088]. The upper bound is $\Lambda(1.49 + \overline{\hat{\mu}_a}) - \Lambda(\overline{\hat{\mu}_a})$ and the lower bound $\Lambda(1.49 + \underline{\hat{\mu}_a}) - \Lambda(\underline{\hat{\mu}_a})$.

The average partial effect of response probability is of primary interest, we calculate the average partial effect of model (4.20). Since the unobserved heterogeneity has rarely, if ever, natural measurements, it is unclear what value we need to plug in for

a. A suggested solution into it is $a_0 = E(a_i) = \alpha_0 + \alpha_1 y_0$. Under Assumption 4.2, the distribution of a_0 is Normal($\alpha_0 + \alpha_1 y_0, \sigma_a^2$) and thus its relevant density function is $f(a_0) = \frac{1}{\sqrt{2\pi\sigma_a}} \exp(-1/2(\frac{a_0 - (\alpha_0 + \alpha_1 y_0)}{\sigma_a})^2)$. The estimated average partial effect of response probability with respective to mean of heterogeneity across i is calculated by the following:

$$\int_{\mathbb{R}} [\Lambda(\hat{\rho} + a) - \Lambda(a)] \hat{f}(a) \, \mathrm{d}\, a, \tag{4.21}$$

where $\hat{f}(a) = \frac{1}{\sqrt{2\pi\hat{\sigma}_a^2}} \exp(-1/2(\frac{a - (\hat{\alpha}_0 + \hat{\alpha}_1 y_0)}{\hat{\sigma}_a})^2)$. It is obvious that the limiting distribution is $\lim_{N\to\infty} \hat{f}(a)$ = f(a). We calculate the integral of (4.21) from $-\infty$ to ∞.

4.5 Conclusion

In this chapter we examine logit model as a specific non-linear case of theoretical framework of conditional maximum likelihood function in estimating non-linear, dynamic, unobserved effects panel data models with feedback (Wooldridge 2000). We make use of a joint density conditional on the strictly exogenous variables and the initial condition. Because we model the density of the unobserved effect conditional on the initial condition and exogenous variables, this is different from the treatment of initial condition as fixed. It is feasible to construct the conditional MLE and the relevant asymptotic property hold as the cross section sample size increase.

The simulation shows that when the true value of ρ is closer to zeros, the estimates away from the true value of parameter is more likely to be rejected and is less likely to be rejected, vice versa when the true value is getting closer to 1. The empirical evidence shows that the unobserved heterogeneity have the negative impact on the willingness to have a union membership. We have considered the important problems of estimating the partial effects at the average value of the unobserved heterogeneity and the partial effects averaged across the distribution of the unobserved heterogeneity in the empirical.

Our approach is parametric method, but the auxiliary conditional density can be arbitrarily modeled in a flexible way leading to a strait forward parameterization that can be estimated using standard software. If we want to extend to consider the exogenous variable, we can use the suggestion of Wooldridge (2000) in which we might carefully specify the conditional density function of $h(a_i|w_i, y_{i0})$, where w_i including exogenous variables and strictly exogenous variables, x_i and z_i. Using the specific case of nonlinear model, this chapter shows that the idea of specifying a conditional distribution for the unobserved heterogeneity given the initial conditions should prove useful for analyzing the dynamic logit model with unobserved effects.

$$\theta = (\rho, 0.2, 0.4, \sqrt{1.2})$$
$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{1.2})$$

P\ $\overset{\rho_0 =}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	ρ
0.01	0.1200	0.0700	0.0360	0.0240	0.0040	0.0080	
0.05	0.2720	0.2040	0.1340	0.0820	0.0500	0.0440	0.25
0.10	0.3980	0.2800	0.2140	0.1440	0.1040	0.0880	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.3	0.35	0.4	0.45	0.5	0.55	ρ
0.01	0.0100	0.0200	0.0460	0.0740	0.1020	0.1760	
0.05	0.0520	0.0800	0.1140	0.1960	0.2740	0.3820	0.25
0.10	0.1060	0.1260	0.2140	0.2840	0.4080	0.5320	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.60	0.65	0.70	0.75	0.80	0.85	ρ
0.01	0.2620	0.3680	0.4960	0.6540	0.7480	0.8320	
0.05	0.5160	0.6720	0.7620	0.8400	0.9080	0.9420	0.25
0.10	0.6800	0.7640	0.8480	0.9100	0.9460	0.9760	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.90	0.95	1.00				ρ
0.01	0.8980	0.9380	0.9700				
0.05	0.9760	0.9800	0.9980				0.25
0.10	0.9800	1.0000	1.0000				

Repetitions=500, $\hat{\theta} = \frac{1}{500}\sum_{j=1}^{500} \theta_j^*$, $\sqrt{1.2} \simeq 1.0954$

(ii)

$$\theta = (\rho, 0.2, 0.4, \sqrt{1.2})$$
$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{1.2})$$

P\ $\overset{\rho_0 =}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	ρ
0.01	0.5860	0.4620	0.3700	0.2400	0.1640	0.1000	
0.05	0.8140	0.7120	0.6020	0.4900	0.3780	0.2680	0.5
0.10	0.8800	0.8300	0.7380	0.6140	0.5160	0.3860	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.3	0.35	0.4	0.45	0.5	0.55	ρ
0.01	0.0540	0.0300	0.0240	0.0100	0.0080	0.0140	
0.05	0.1760	0.1140	0.0660	0.0500	0.0480	0.0540	0.5
0.10	0.2760	0.1860	0.1320	0.0880	0.0880	0.1020	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.6	0.65	0.7	0.75	0.8	0.85	ρ
0.01	0.0240	0.0400	0.0640	0.1080	0.1640	0.2420	
0.05	0.0780	0.1240	0.1820	0.2840	0.3960	0.5060	0.5
0.10	0.1460	0.1940	0.2940	0.4060	0.5300	0.6360	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.9	0.95	1				ρ
0.01	0.3820	0.4720	0.6120				
0.05	0.6280	0.7540	0.8400				0.5
0.10	0.7640	0.8420	0.9000				

Repetitions=500, $\hat{\theta} = \frac{1}{500}\sum_{j=1}^{500}\theta_j^*$, $\sqrt{1.2} \simeq 1.0954$

(iii)

$$\theta = (\rho, 0.2, 0.4, \sqrt{1.2})$$

$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{1.2})$$

P\ $\overset{\rho_0 =}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	ρ
0.01	0.9280	0.8820	0.8360	0.7620	0.6640	0.5700	
0.05	0.9760	0.9620	0.9340	0.8960	0.8500	0.7800	0.75
0.10	0.9920	0.9780	0.9700	0.9420	0.9020	0.8540	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.3	0.35	0.4	0.45	0.5	0.55	ρ
0.01	0.4500	0.3500	0.2480	0.1640	0.0940	0.0440	
0.05	0.6940	0.5840	0.4860	0.3860	0.2680	0.1760	0.75
0.10	0.7880	0.7100	0.6040	0.5000	0.3960	0.2860	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.6	0.65	0.7	0.75	0.8	0.85	ρ
0.01	0.0240	0.0140	0.0080	0.0100	0.0120	0.0240	
0.05	0.1100	0.0600	0.0440	0.0420	0.0480	0.0780	0.75
0.10	0.1940	0.1360	0.0900	0.0820	0.1000	0.1440	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.9	0.95	1				ρ
0.01	0.0300	0.0600	0.1080				
0.05	0.1340	0.1820	0.2620				0.75
0.10	0.2060	0.2780	0.3800				

Repetitions=500, $\hat{\theta} = \frac{1}{500}\sum_{j=1}^{500} \theta_j^*$, $\sqrt{1.2} \simeq 1.0954$

(iv)

$$\theta = (\rho, 0.2, 0.4, \sqrt{1.2})$$

$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{1.2})$$

P\ $\overset{\rho_0 =}{\rightarrow}$	0.00	0.05	0.10	0.15	0.20	0.25	ρ
0.01	0.9780	0.9600	0.9300	0.8940	0.8520	0.8040	
0.05	0.9960	0.9940	0.9820	0.9660	0.9360	0.9040	0.9
0.10	0.9980	0.9960	0.9940	0.9880	0.9720	0.9400	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.30	0.35	0.40	0.45	0.50	0.55	ρ
0.01	0.7360	0.6540	0.5480	0.4340	0.3120	0.2400	
0.05	0.8700	0.8220	0.7540	0.6900	0.5940	0.4700	0.9
0.10	0.9140	0.8780	0.8300	0.7600	0.7000	0.6160	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.6	0.65	0.7	0.75	0.8	0.85	ρ
0.01	0.1680	0.0880	0.0540	0.0180	0.0160	0.0120	
0.05	0.3500	0.2720	0.1800	0.1000	0.0620	0.0400	0.9
0.10	0.4900	0.3640	0.2820	0.1920	0.1220	0.1020	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.9	0.95	1				ρ
0.01	0.0100	0.0140	0.0300				
0.05	0.0540	0.0680	0.1040				0.9
0.10	0.0960	0.1180	0.1620				

Repetitions=500, $\hat{\theta} = \frac{1}{500}\sum_{j=1}^{500} \theta_j^*$, $\sqrt{1.2} \simeq 1.0954$

(v)

$$\theta = (\rho, 0.2, 0.4, \sqrt{1.2})$$
$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{1.2})$$

P\ $\overset{\rho_0 =}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	ρ
0.01	0.9700	0.9600	0.9500	0.9100	0.9000	0.8300	
0.05	1.0000	1.0000	0.9800	0.9600	0.9600	0.9300	0.95
0.10	1.0000	1.0000	1.0000	0.9900	0.9700	0.9600	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.3	0.35	0.4	0.45	0.5	0.55	ρ
0.01	0.7700	0.7100	0.6300	0.4800	0.3000	0.2000	
0.05	0.9000	0.8500	0.7900	0.7300	0.6500	0.5300	0.95
0.10	0.9400	0.9000	0.8600	0.7900	0.7300	0.6700	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.6	0.65	0.7	0.75	0.8	0.85	ρ
0.01	0.1600	0.1000	0.0500	0.0100	0.0100	0.0100	
0.05	0.3400	0.2400	0.1700	0.1000	0.0600	0.0200	0.95
0.10	0.5800	0.3800	0.2800	0.1800	0.1200	0.1000	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.9	0.95	1				ρ
0.01	0.0100	0.0100	0.0200				
0.05	0.0500	0.0600	0.0900				0.95
0.10	0.0700	0.0900	0.1100				

Repetitions=500, $\hat{\theta} = \frac{1}{500}\sum_{j=1}^{500}\theta_j^*$, $\sqrt{1.2} \simeq 1.0954$

(vi)

$$\theta = (\rho, 0.2, 0.4, \sqrt{1.25})$$

$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{1.25})$$

P\ $\overset{\rho_0 =}{\longrightarrow}$	0	0.05	0.1	0.15	0.2	0.25	ρ
0.01	0.1260	0.0800	0.0300	0.0140	0.0040	0.0000	
0.05	0.3020	0.1900	0.1320	0.0860	0.0340	0.0280	0.25
0.10	0.3960	0.3100	0.1980	0.1420	0.1040	0.0720	
P\ $\overset{\rho_0 =}{\longrightarrow}$	0.3	0.35	0.4	0.45	0.5	0.55	ρ
0.01	0.0000	0.0140	0.0380	0.0700	0.1140	0.2200	
0.05	0.0400	0.0760	0.1320	0.2280	0.3360	0.4400	0.25
0.10	0.0900	0.1460	0.2420	0.3440	0.4460	0.5620	
P\ $\overset{\rho_0 =}{\longrightarrow}$	0.60	0.65	0.70	0.75	0.80	0.85	ρ
0.01	0.3120	0.4260	0.5220	0.6460	0.7500	0.8340	
0.05	0.5440	0.6580	0.7680	0.8400	0.9040	0.9480	0.25
0.10	0.6680	0.7760	0.8460	0.9060	0.9560	0.9760	
P\ $\overset{\rho_0 =}{\longrightarrow}$	0.90	0.95	1.00				ρ
0.01	0.8960	0.9420	0.9700				
0.05	0.9720	0.9920	1.0000				0.25
0.10	0.9920	1.0000	1.0000				

Repetitions=500, $\hat{\theta} = \frac{1}{500}\sum_{j=1}^{500}\theta_j^*$, $\sqrt{1.25} \simeq 1.1180$

(ii)

$$\theta = (\rho, 0.2, 0.4, \sqrt{1.25})$$

$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{1.25})$$

P\ $\overset{\rho_0 =}{\to}$	0	0.05	0.1	0.15	0.2	0.25	ρ
0.01	0.5560	0.4760	0.3820	0.2660	0.1720	0.1140	
0.05	0.7920	0.6740	0.5740	0.4920	0.4020	0.2800	0.5
0.10	0.8680	0.8000	0.6840	0.5960	0.4980	0.4120	
P\ $\overset{\rho_0 =}{\to}$	0.3	0.35	0.4	0.45	0.5	0.55	ρ
0.01	0.0580	0.0280	0.0140	0.0080	0.0040	0.0140	
0.05	0.1900	0.1240	0.0740	0.0460	0.0400	0.0460	0.5
0.10	0.3000	0.2100	0.1460	0.1000	0.0700	0.1060	
P\ $\overset{\rho_0 =}{\to}$	0.6	0.65	0.7	0.75	0.8	0.85	ρ
0.01	0.0200	0.0360	0.0680	0.1160	0.1860	0.2940	
0.05	0.0820	0.1340	0.2040	0.3160	0.4260	0.5200	0.5
0.10	0.1520	0.2260	0.3300	0.4420	0.5320	0.6200	
P\ $\overset{\rho_0 =}{\to}$	0.9	0.95	1				ρ
0.01	0.4060	0.5040	0.5940				
0.05	0.6080	0.7240	0.8240				0.5
0.10	0.7440	0.8320	0.8980				

Repetitions=500, $\hat{\theta} = \frac{1}{500}\sum_{j=1}^{500} \theta_j^*$, $\sqrt{1.25} \simeq 1.1180$

(iii)

$$\theta = (\rho, 0.2, 0.4, \sqrt{1.25})$$
$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{1.25})$$

P\ $\overset{\rho_0 =}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	ρ
0.01	0.9320	0.8820	0.8100	0.7340	0.6240	0.5380	
0.05	0.9820	0.9660	0.9400	0.9000	0.8360	0.7500	0.75
0.10	0.9880	0.9820	0.9700	0.9440	0.9080	0.8480	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.3	0.35	0.4	0.45	0.5	0.55	ρ
0.01	0.4300	0.3240	0.2500	0.1400	0.0940	0.0560	
0.05	0.6560	0.5700	0.4520	0.3620	0.2740	0.1620	0.75
0.10	0.7640	0.6700	0.5760	0.4680	0.3800	0.2760	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.6	0.65	0.7	0.75	0.8	0.85	ρ
0.01	0.0260	0.1060	0.0040	0.0060	0.0100	0.0200	
0.05	0.1060	0.0700	0.0440	0.0400	0.0400	0.0700	0.75
0.10	0.1840	0.1280	0.1040	0.0760	0.1040	0.1500	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.9	0.95	1				ρ
0.01	0.0320	0.0600	0.1200				
0.05	0.1340	0.2060	0.3040				0.75
0.10	0.2160	0.3200	0.4140				

Repetitions=500, $\hat{\theta} = \frac{1}{500}\sum_{j=1}^{500}\theta_j^*$, $\sqrt{1.25} \simeq 1.1180$

iv

$$\theta = (\rho, 0.2, 0.4, \sqrt{1.25})$$

$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{1.25})$$

P\ $\overset{\rho_0 =}{\rightarrow}$	0.00	0.05	0.10	0.15	0.20	0.25	ρ
0.01	0.9860	0.9760	0.9600	0.9260	0.8720	0.8220	
0.05	0.9940	0.9940	0.9880	0.9800	0.9640	0.9360	0.9
0.10	0.9980	0.9960	0.9940	0.9880	0.9820	0.9680	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.30	0.35	0.40	0.45	0.50	0.55	ρ
0.01	0.7160	0.6040	0.5220	0.4140	0.3180	0.2240	
0.05	0.8960	0.8360	0.7480	0.6420	0.5460	0.4480	0.9
0.10	0.9420	0.9120	0.8440	0.7700	0.6640	0.5620	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.6	0.65	0.7	0.75	0.8	0.85	ρ
0.01	0.1600	0.0940	0.0460	0.0240	0.0140	0.0120	
0.05	0.3520	0.2500	0.1740	0.1180	0.0720	0.0420	0.9
0.10	0.4700	0.3740	0.2680	0.1960	0.1400	0.0920	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.9	0.95	1				ρ
0.01	0.0100	0.0120	0.0180				
0.05	0.0340	0.0480	0.0740				0.9
0.10	0.0760	0.0940	0.1380				

Repetitions=500, $\hat{\theta} = \frac{1}{500}\sum_{j=1}^{500} \theta_j^*$, $\sqrt{1.25} \simeq 1.1180$

(v)

$$\theta = (\rho, 0.2, 0.4, \sqrt{1.25})$$
$$\theta_0 = (\rho_0, 0.2, 0.4, \sqrt{1.25})$$

P\ $\rho_0 \overset{=}{\to}$	0	0.05	0.1	0.15	0.2	0.25	ρ
0.01	0.9920	0.9840	0.9700	0.9500	0.9300	0.8720	
0.05	0.9960	0.9960	0.9940	0.9860	0.9740	0.9640	0.95
0.10	1.0000	0.9960	0.9960	0.9940	0.9880	0.9760	
P\ $\rho_0 \overset{=}{\to}$	0.3	0.35	0.4	0.45	0.5	0.55	ρ
0.01	0.8040	0.7120	0.5980	0.5120	0.4040	0.3020	
0.05	0.9340	0.8960	0.8260	0.7460	0.6320	0.5340	0.95
0.10	0.9640	0.9420	0.9060	0.8400	0.7660	0.6540	
P\ $\rho_0 \overset{=}{\to}$	0.6	0.65	0.7	0.75	0.8	0.85	ρ
0.01	0.2200	0.1460	0.0920	0.0460	0.0200	0.0140	
0.05	0.4400	0.3280	0.2440	0.1660	0.1080	0.0620	0.95
0.10	0.5540	0.4560	0.3580	0.2640	0.1860	0.1300	
P\ $\rho_0 \overset{=}{\to}$	0.9	0.95	1				ρ
0.01	0.0160	0.0100	0.0100				
0.05	0.0340	0.0400	0.0520				0.95
0.10	0.0900	0.0700	0.0860				

Repetitions=500, $\hat{\theta} = \frac{1}{500}\sum_{j=1}^{500} \theta_j^*$, $\sqrt{1.25} \simeq 1.1180$

(vi)

$$\theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{1.2})$$

$$\theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{1.2})$$

P\ $\overset{\rho_0=}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	ρ
0.01	0.5833	0.4467	0.3300	0.2167	0.1433	0.0767	
0.05	0.8267	0.6967	0.6033	0.4867	0.3433	0.2367	0.5
0.10	0.8967	0.8333	0.7133	0.6167	0.4967	0.3633	
P\ $\overset{\rho_0=}{\rightarrow}$	0.3	0.35	0.4	0.45	0.5	0.55	ρ
0.01	0.0400	0.0233	0.0067	0.0067	0.0033	0.0100	
0.05	0.1500	0.0867	0.0467	0.0367	0.0333	0.0367	0.5
0.10	0.2433	0.1567	0.1033	0.0733	0.0633	0.0767	
P\ $\overset{\rho_0=}{\rightarrow}$	0.60	0.65	0.70	0.75	0.80	0.85	ρ
0.01	0.0167	0.0267	0.0467	0.0800	0.1567	0.2467	
0.05	0.0567	0.1000	0.1700	0.2767	0.4033	0.5233	0.5
0.10	0.1033	0.1800	0.2900	0.4167	0.5467	0.6833	
P\ $\overset{\rho_0=}{\rightarrow}$	0.90	0.95	1.00				ρ
0.01	0.3833	0.5100	0.6367				
0.05	0.6600	0.7700	0.8533				0.5
0.10	0.7800	0.8600	0.9167				

Repetitions=300, $\hat{\theta} = \frac{1}{300}\sum_{j=1}^{300}\theta_j^*$, $\sqrt{1.2} \simeq 1.0954$

(ii)

$$\theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{1.2})$$
$$\theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{1.2})$$

P\ $\overset{\rho_0=}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	ρ
0.01	0.9867	0.9767	0.9467	0.9167	0.8667	0.8233	
0.05	0.9967	0.9933	0.9900	0.9800	0.9567	0.9267	0.9
0.10	0.9967	0.9967	0.9933	0.9900	0.9800	0.9700	
P\ $\overset{\rho_0=}{\rightarrow}$	0.3	0.35	0.4	0.45	0.5	0.55	ρ
0.01	0.7533	0.6400	0.5267	0.4033	0.3067	0.2100	
0.05	0.8733	0.8533	0.7733	0.6800	0.5600	0.4533	0.9
0.10	0.9333	0.8900	0.8600	0.7967	0.6967	0.5833	
P\ $\overset{\rho_0=}{\rightarrow}$	0.6	0.65	0.7	0.75	0.8	0.85	ρ
0.01	0.1300	0.0767	0.0267	0.0133	0.0033	0.0067	
0.05	0.3167	0.2333	0.1500	0.0967	0.0467	0.0300	0.9
0.10	0.4833	0.3467	0.2700	0.1767	0.1067	0.0700	
P\ $\overset{\rho_0=}{\rightarrow}$	0.9	0.95	1				ρ
0.01	0.0067	0.0067	0.0133				
0.05	0.0233	0.0433	0.0833				0.9
0.10	0.0633	0.0967	0.1300				

Repetitions=300, $\hat{\theta} = \frac{1}{300}\sum_{j=1}^{300}\theta_j^*$, $\sqrt{1.2} \simeq 1.0954$

(iii)

$$\theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{1.25})$$

$$\theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{1.25})$$

P\ $\overset{\rho_0 =}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	ρ
0.01	0.5533	0.4500	0.3533	0.2667	0.1767	0.1100	
0.05	0.7533	0.6567	0.5700	0.4800	0.3733	0.2867	0.5
0.10	0.8467	0.7600	0.6600	0.5800	0.4967	0.3833	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.3	0.35	0.4	0.45	0.5	0.55	ρ
0.01	0.0533	0.0267	0.0133	0.0067	0.0033	0.0033	
0.05	0.1900	0.1167	0.0633	0.0300	0.0333	0.0533	0.5
0.10	0.2933	0.2000	0.1300	0.0933	0.0833	0.1167	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.60	0.65	0.70	0.75	0.80	0.85	ρ
0.01	0.0100	0.0367	0.0767	0.1367	0.2267	0.3233	
0.05	0.0867	0.1500	0.2467	0.3367	0.4233	0.5333	0.5
0.10	0.1667	0.2500	0.3567	0.4433	0.5400	0.6600	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.90	0.95	1.00				ρ
0.01	0.4067	0.5233	0.6133				
0.05	0.6367	0.7100	0.8133				0.5
0.10	0.7233	0.8267	0.8993				

Repetitions=300, $\hat{\theta} = \frac{1}{300}\sum_{j=1}^{300} \theta_j^*$, $\sqrt{1.25} \simeq 1.1180$

(ii)

$$\theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{1.25})$$

$$\theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{1.25})$$

P\ $\rho_0 = \rightarrow$	0	0.05	0.1	0.15	0.2	0.25	ρ
0.01	0.9933	0.9867	0.9767	0.9333	0.8733	0.7867	
0.05	1.0000	1.0000	0.9967	0.9933	0.9800	0.9467	0.9
0.10	1.0000	1.0000	1.0000	0.9967	0.9933	0.9867	
P\ $\rho_0 = \rightarrow$	0.3	0.35	0.4	0.45	0.5	0.55	ρ
0.01	0.7067	0.6100	0.5233	0.4033	0.2933	0.2133	
0.05	0.8900	0.8233	0.7200	0.6500	0.5667	0.4467	0.9
0.10	0.9500	0.9067	0.8300	0.7333	0.6633	0.5733	
P\ $\rho_0 = \rightarrow$	0.6	0.65	0.7	0.75	0.8	0.85	ρ
0.01	0.1433	0.0900	0.0600	0.0367	0.0200	0.0067	
0.05	0.3500	0.2433	0.1600	0.1033	0.0733	0.0433	0.9
0.10	0.4633	0.3600	0.2567	0.1733	0.1167	0.0900	
P\ $\rho_0 = \rightarrow$	0.9	0.95	1				ρ
0.01	0.0067	0.0000	0.0067				
0.05	0.0267	0.0267	0.0533				0.9
0.10	0.0667	0.0900	0.1467				

Repetitions=300, $\hat{\theta} = \frac{1}{300}\sum_{j=1}^{300} \theta_j^*$, $\sqrt{1.25} \simeq 1.1180$

(iii)

Table 4.1: Model a :$H_0 : \theta = \theta_0$, where ρ =0 ~ 0.95

$$\begin{aligned} \theta &= (\rho, 0.2, 0.4, \sqrt{1.2}) \\ \theta_0 &= (\rho_0, 0.2, 0.4, \sqrt{1.2}) \end{aligned}$$

P\ $\hat{\theta}$	$\hat{\rho}$ (1.2×10^{-3})	$\hat{\alpha_0}$ (0.1998)	$\hat{\alpha_1}$ (0.4053)	$\hat{\sigma_a}$ (1.0933)	ρ_0
0.01	0.0040	0.0080	0.0120	0.0040	
0.05	0.0460	0.0260	0.0460	0.0540	0.00
0.10	0.0900	0.0600	0.1060	0.1020	
P\ $\hat{\theta}$	(0.2453)	(0.2036)	(0.4049)	(1.0916)	ρ_0
0.01	0.0080	0.0080	0.0080	0.0140	
0.05	0.0440	0.0400	0.0560	0.0420	0.25
0.10	0.0880	0.0760	0.1120	0.0840	
$P\backslash\hat{\theta}$	(0.4918)	(0.2068)	(0.4046)	(1.0925)	ρ_0
0.01	0.0080	0.0120	0.0120	0.0100	
0.05	0.0480	0.0400	0.0460	0.0520	0.5
0.10	0.0880	0.0820	0.1040	0.1060	
$P\backslash\hat{\theta}$	(0.7442)	(0.2073)	(0.4067)	(1.0937)	ρ_0
0.01	0.0100	0.0120	0.0140	0.0040	
0.05	0.0420	0.0380	0.0440	0.0520	0.75
0.10	0.0820	0.0900	0.1220	0.1060	
P\ $\hat{\theta}$	(0.8915)	(0.2087)	(0.4097)	(1.0967)	ρ_0
0.01	0.0100	0.0100	0.0120	0.0020	
0.05	0.0540	0.0480	0.0600	0.5500	0.9
0.10	0.0960	0.0980	0.1060	0.8500	
P\ $\hat{\theta}$	(0.9413)	(0.2093)	(0.4096)	(1.0986)	ρ_0
0.01	0.0100	0.0200	0.0100	0.0100	
0.05	0.0600	0.0600	0.0800	0.0300	0.95
0.10	0.0900	0.0800	0.1700	0.0800	

Repetitions=500, $\hat{\theta} = \frac{1}{500}\sum_{j=1}^{500} \theta_j^*$, $\sqrt{1.2} \simeq 1.0954$

Table 4.2: Model a: $H_0 : \rho = \rho_0$, where ρ =0 $\sim$ 0.95

$$\begin{aligned} \theta &= (\rho, 0.2, 0.4, \sqrt{1.2}) \\ \theta_0 &= (\rho_0, 0.2, 0.4, \sqrt{1.2}) \end{aligned}$$

P\ $\overset{\rho_0 =}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	ρ
0.01	0.0040	0.0100	0.0200	0.0380	0.0680	0.1160	
0.05	0.0460	0.0460	0.0720	0.1180	0.1800	0.2760	0
0.10	0.0900	0.1020	0.1240	0.2040	0.2860	0.4200	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.30	0.35	0.40	0.45	0.50	0.55	ρ
0.01	0.1740	0.2720	0.4040	0.4920	0.6630	0.7600	
0.05	0.4120	0.5000	0.6540	0.7680	0.8560	0.9220	0
0.10	0.5060	0.6460	0.7760	0.8860	0.9260	0.9480	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.60	0.65	0.70	0.75	0.80	0.85	ρ
0.01	0.8580	0.9160	0.9480	0.9740	0.9940	0.9980	
0.05	0.9480	0.9740	0.9940	0.9880	1.0000	1.0000	0
0.10	0.9740	0.9940	1.0000	1.0000	1.0000	1.0000	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.90	0.95	1.000				ρ
0.01	1.0000	1.0000	1.0000				
0.05	1.0000	1.0000	1.0000				0
0.10	1.0000	1.0000	1.0000				

Repetitions=500, $\hat{\theta} = \frac{1}{500}\sum_{j=1}^{500}\theta_j^*$, $\sqrt{1.2} \simeq 1.0954$

(i)

Table 4.3: Model b :$H_0 : \theta = \theta_0$, where ρ =0 ~ 0.95

$$\begin{aligned}\theta &= (\rho, 0.2, 0.4, \sqrt{1.25})\\ \theta_0 &= (\rho_0, 0.2, 0.4, \sqrt{1.25})\end{aligned}$$

$P\backslash\hat{\theta}$	$\hat{\rho}$ (-6.5×10^{-3})	$\hat{\alpha_0}$ (0.2253)	$\hat{\alpha_1}$ (0.3803)	$\hat{\sigma_a}$ (1.0699)	ρ_0
0.01	0.0040	0.0060	0.0120	0.0200	
0.05	0.0400	0.0440	0.0460	0.0760	0.00
0.10	0.0980	0.0960	0.0940	0.1540	
$P\backslash\hat{\theta}$	(0.2392)	(0.2253)	(0.3878)	(1.0733)	ρ_0
0.01	0.0000	0.0080	0.0120	0.0200	
0.05	0.0280	0.0400	0.0480	0.0800	0.25
0.10	0.0720	0.0800	0.0780	0.1280	
$P\backslash\hat{\theta}$	(0.4894)	(0.2284)	(0.3872)	(1.0741)	ρ_0
0.01	0.0040	0.0080	0.0140	0.0180	
0.05	0.0400	0.0340	0.0480	0.0740	0.5
0.10	0.0700	0.0880	0.0860	0.1400	
$P\backslash\hat{\theta}$	(0.7357)	(0.2326)	(0.3874)	(1.0778)	ρ_0
0.01	0.0060	0.0080	0.0100	0.0180	
0.05	0.0400	0.0420	0.0520	0.0660	0.75
0.10	0.0760	0.0900	0.0940	0.1220	
$P\backslash\hat{\theta}$	(0.8900)	(0.2304)	(0.3872)	(1.0770)	ρ_0
0.01	0.0100	0.0100	0.0060	0.0200	
0.05	0.0340	0.0420	0.0460	0.0780	0.9
0.10	0.0760	0.0920	0.0900	0.1200	
$P\backslash\hat{\theta}$	(0.9379)	(0.2322)	(0.3905)	(1.0794)	ρ_0
0.01	0.0100	0.0080	0.0040	0.0160	
0.05	0.0400	0.0440	0.0480	0.0820	0.95
0.10	0.0700	0.0900	0.0940	0.1220	

Repetitions=500, $\hat{\theta} = \frac{1}{500}\sum_{j=1}^{500}\theta_j^*$, $\sqrt{1.25} \simeq 1.1180$

Table 4.4: Model b: $H_0 : \rho = \rho_0$, where ρ =0 $\sim$ 0.95

$$\begin{aligned} \theta &= (\rho, 0.2, 0.4, \sqrt{1.25}) \\ \theta_0 &= (\rho_0, 0.2, 0.4, \sqrt{1.25}) \end{aligned}$$

P\ $\overset{\rho_0 =}{\rightarrow}$	0	0.05	0.1	0.15	0.2	0.25	ρ
0.01	0.0040	0.0080	0.0160	0.0440	0.0820	0.1260	
0.05	0.0400	0.0620	0.0880	0.1340	0.2200	0.3200	0
0.10	0.0980	0.1180	0.1520	0.2340	0.3220	0.4680	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.30	0.35	0.40	0.45	0.50	0.55	ρ
0.01	0.2140	0.3100	0.4480	0.5340	0.6380	0.7420	
0.05	0.4640	0.5440	0.6520	0.7520	0.8460	0.9040	0
0.10	0.5520	0.6720	0.7640	0.8520	0.9100	0.9540	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.60	0.65	0.70	0.75	0.80	0.85	ρ
0.01	0.8360	0.8980	0.9520	0.9700	0.9920	0.9960	
0.05	0.9540	0.9720	0.9920	0.9960	1.0000	1.0000	0
0.10	0.9720	0.9920	0.9960	1.0000	1.0000	1.0000	
P\ $\overset{\rho_0 =}{\rightarrow}$	0.90	0.95	1.000				ρ
0.01	1.0000	1.0000	1.0000				
0.05	1.0000	1.0000	1.0000				0
0.10	1.0000	1.0000	1.0000				

Repetitions=500, $\hat{\theta} = \frac{1}{500}\sum_{j=1}^{500} \theta_j^*$, $\sqrt{1.25} \simeq 1.1180$

(i)

Table 4.5: Model c :$H_0 : \theta = \theta_0$, where ρ =0 $\sim$ 0.9

$$\begin{aligned}\theta &= (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{1.2})\\ \theta_0 &= (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{1.2})\end{aligned}$$

P\\$\hat{\theta}$	$\hat{\rho}$ (-0.0108)	$\hat{\beta}$ (0.1449)	$\hat{\alpha}_0$ (0.2085)	$\hat{\alpha}_1$ (0.4093)	$\hat{\alpha}_2$ (0.3741)	$\hat{\sigma}_a$ (1.0886)	ρ_0
0.01	0.0000	0.0000	0.0100	0.0233	0.0100	0.0067	
0.05	0.0267	0.0667	0.0367	0.0633	0.0600	0.0200	0.00
0.10	0.0900	0.0833	0.0867	0.1267	0.1200	0.0567	
P\\$\hat{\theta}$	(0.4884)	(0.1466)	(0.2086)	(0.4140)	(0.3717)	(1.0876)	ρ_0
0.01	0.0033	0.0100	0.0100	0.0133	0.0133	0.0033	
0.05	0.0333	0.0300	0.0400	0.0700	0.0600	0.0333	0.5
0.10	0.0633	0.0833	0.0733	0.1300	0.1133	0.0600	
P\\$\hat{\theta}$	(0.8901)	(0.1471)	(0.2129)	(0.4141)	(0.3670)	(0.10910)	ρ_0
0.01	0.0067	0.0067	0.0100	0.0200	0.0167	0.0033	
0.05	0.0233	0.0500	0.0433	0.0767	0.0433	0.0333	0.9
0.10	0.0633	0.0833	0.0900	0.1167	0.1000	0.0067	

Repetitions=300, $\hat{\theta} = \frac{1}{300}\sum_{j=1}^{300}\theta_j^*$, $\sqrt{1.2} \simeq 1.0954$

a_i : Normal

Table 4.6: Model c: $H_0 : \rho = \rho_0$, where ρ =0 $\sim$ 0.9

$$\begin{aligned} \theta &= (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{1.2}) \\ \theta_0 &= (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{1.2}) \end{aligned}$$

P\ $\overset{\rho_0 =}{\longrightarrow}$	0	0.05	0.1	0.15	0.2	0.25	ρ
0.01	0.0000	0.0067	0.0100	0.0400	0.0567	0.1067	
0.05	0.0267	0.0433	0.0633	0.1200	0.1900	0.2933	0
0.10	0.0900	0.0933	0.1400	0.1933	0.3033	0.4500	
P\ $\overset{\rho_0 =}{\longrightarrow}$	0.30	0.35	0.40	0.45	0.50	0.55	ρ
0.01	0.1900	0.2700	0.4167	0.5533	0.6767	0.7967	
0.05	0.4333	0.5833	0.7000	0.8067	0.8600	0.9100	0
0.10	0.5933	0.7033	0.8100	0.8667	0.9100	0.9533	
P\ $\overset{\rho_0 =}{\longrightarrow}$	0.60	0.65	0.70	0.75	0.80	0.85	ρ
0.01	0.0.8600	0.9000	0.9500	0.9767	0.9933	0.9967	
0.05	0.9533	0.9767	0.9967	0.9967	1.0000	1.0000	0
0.10	0.9833	0.9967	1.0000	1.0000	1.0000	1.0000	
P\ $\overset{\rho_0 =}{\longrightarrow}$	0.90	0.95	1.000				ρ
0.01	1.0000	1.0000	1.0000				
0.05	1.0000	1.0000	1.0000				0
0.10	1.0000	1.0000	1.0000				

Repetitions=300, $\hat{\theta} = \frac{1}{300} \sum_{j=1}^{300} \theta_j^*$, $\sqrt{1.2} \simeq 1.0954$

a_i : Normal

(i)

Table 4.7: Model d : $H_0 : \theta = \theta_0$, where ρ =0 $\sim$ 0.9

$$\theta = (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{1.25})$$
$$\theta_0 = (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{1.25})$$

$P \backslash \hat{\theta}$	$\hat{\rho}$ (-0.0075)	$\hat{\beta}$ (0.1454)	$\hat{\alpha}_0$ (0.2225)	$\hat{\alpha}_1$ (0.3845)	$\hat{\alpha}_2$ (0.3777)	$\hat{\sigma}_a$ (1.0678)	ρ_0
0.01	0.0000	0.0000	0.0033	0.0067	0.0067	0.0167	
0.05	0.0267	0.0700	0.0433	0.0500	0.0400	0.0700	0.00
0.10	0.0767	0.1133	0.0867	0.0967	0.0967	0.1433	
$P \backslash \hat{\theta}$	(0.4837)	(0.1425)	(0.2275)	(0.3909)	(0.3871)	(1.0790)	ρ_0
0.01	0.0033	0.0533	0.0333	0.0500	0.0333	0.0533	
0.05	0.0333	0.0533	0.0333	0.0500	0.0333	0.0533	0.5
0.10	0.0833	0.1233	0.0733	0.0933	0.0967	0.1067	
$P \backslash \hat{\theta}$	(0.8909)	(0.1441)	(0.2307)	(0.3825)	(0.3779)	(1.0789)	ρ_0
0.01	0.0067	0.0200	0.0033	0.0233	0.0033	0.0133	
0.05	0.0267	0.0667	0.0267	0.0433	0.0367	0.0433	0.9
0.10	0.0067	0.1167	0.0767	0.1033	0.0700	0.0900	

Repetitions=300, $\hat{\theta} = \frac{1}{300}\sum_{j=1}^{300}\theta_j^*$, $\sqrt{1.25} \simeq 1.1180$

a_i : Non-normal

Table 4.8: Model d: $H_0 : \rho = \rho_0$, where ρ =0 ~ 0.9

$$\begin{aligned}\theta &= (\rho, 0.15, 0.2, 0.4, 0.35, \sqrt{1.25})\\ \theta_0 &= (\rho_0, 0.15, 0.2, 0.4, 0.35, \sqrt{1.25})\end{aligned}$$

P\ $\overset{\rho_0 =}{\longrightarrow}$	0	0.05	0.1	0.15	0.2	0.25	ρ
0.01	0.0000	0.0133	0.0167	0.0333	0.0667	0.1167	
0.05	0.0267	0.0400	0.0767	0.1333	0.2100	0.3267	0
0.10	0.0767	0.0900	0.1400	0.2200	0.3400	0.4600	
P\ $\overset{\rho_0 =}{\longrightarrow}$	0.30	0.35	0.40	0.45	0.50	0.55	ρ
0.01	0.2000	0.3033	0.4367	0.5233	0.6500	0.7600	
0.05	0.4567	0.5400	0.6600	0.7667	0.8700	0.9100	0
0.10	0.5433	0.6667	0.7767	0.8733	0.9133	0.9600	
P\ $\overset{\rho_0 =}{\longrightarrow}$	0.60	0.65	0.70	0.75	0.80	0.85	ρ
0.01	0.8600	0.9100	0.9467	0.9833	0.9933	0.9967	
0.05	0.9467	0.9867	0.9933	0.9967	1.0000	1.0000	0
0.10	0.9867	0.9933	0.9967	1.0000	1.0000	1.0000	
P\ $\overset{\rho_0 =}{\longrightarrow}$	0.90	0.95	1.000				ρ
0.01	1.0000	1.0000	1.0000				
0.05	1.0000	1.0000	1.0000				0
0.10	1.0000	1.0000	1.0000				

Repetitions=300, $\hat{\theta} = \frac{1}{300}\sum_{j=1}^{300} \theta_j^*$, $\sqrt{1.25} \simeq 1.180$

a_i : Non-normal

(i)

Table 4.9: Empirical Evidence for Labor Union Membership, Period:1980 ~ 1987

coefficient	$\hat{\rho}$	$\hat{\alpha_0}$	$\hat{\alpha_1}$	$\hat{\sigma}_a$
CMLE	1.4923	-3.27750	2.6690	1.9997
t-statistics	(9.498)	(18.942)	(8.993)	(12.036)

Appendix A

Conditional mean and variance

The appendix A is to slove the $E(y_i|y_{i0})_{T\times 1}$ and $\mathrm{Var}(y_i|y_{i0})_{T\times T}$ of (2.19). The regression equation (2.1) can be rewritten in terms of y_{i0} and the errors as $y_{it} = \rho^t y_{i0} + \frac{1-\rho^t}{1-\rho} a_i + \sum_{j=1}^{t} \rho^{j-1} \varepsilon_{i,t-j+1}$. According to assumptions on ε_{it} and a_i in Chapter 2, we have $E(a_i|y_{i0}) = \alpha_0 + \alpha_1 y_{i0}$ and hence $E(y_{it}|y_{i0}) = \rho^t y_{i0} + \frac{1-\rho^t}{1-\rho}(\alpha_1 + \rho y_{i0})$ for t=1,. . .,T. We can write the expectation of T observaitons on individual i conditional on y_{i0} as follows:

$$E(y_i|y_{i0}) = \begin{pmatrix} \alpha_0 + (\alpha_1 + \rho) y_{i0} & \cdots & \frac{1-\rho^t}{1-\rho}\alpha_0 + (\frac{1-\rho^t}{1-\rho}\alpha_1 + \rho^t) y_{i0} & \cdots \end{pmatrix}_{T\times 1},$$

where $t = 1, \ldots, T$.

$$\begin{aligned} \Omega(y_{i0}) &\equiv Var(y_i|y_{i0}) \\ &= E\left((y_i - E(y_i|y_{i0}))(y_i - E(y_i|y_{i0}))'|y_{i0}\right) \\ &= E(y_i y_i'|y_{i0}) - E(y_i|y_{i0})E(y_i|y_{i0})' \end{aligned}$$

$$\Omega(y_{i0}) = \begin{pmatrix} \omega_{11} & \ldots & \omega_{1T} \\ \vdots & \ddots & \vdots \\ \omega_{T1} & \ldots & \omega_{TT} \end{pmatrix},$$

where $\omega_{tt} = E(y_{it}^2|y_{i0}) - E(y_{it}|y_{i0})^2$ and $\omega_{st} = E(y_{is}y_{it}|y_{i0}) - E(y_{is}|y_{i0})E(y_{it}|y_{i0})$. To obtain all the elements of the covariance matrix, we need to solve the $E(y_{it}^2|y_{i0})$ and

$E(y_{is}y_{it}|y_{i0})$ for t, $s = 1,\ldots,T$. We write the results as follows:

$$\begin{aligned}\omega_{tt} &= (\frac{1-\rho^t}{1-\rho})^2\sigma_a^2 \;+\; (\frac{1-\rho^{2t}}{1-\rho^2})\sigma_\varepsilon^2,\ t=1,\ldots,T,\\ \omega_{st} &= (\frac{1-\rho^s}{1-\rho}\frac{1-\rho^t}{1-\rho})\sigma_a^2 + \rho^{|t-s|}(\frac{1-\rho^{2s}}{1-\rho^2})\sigma_\varepsilon^2,\ s\neq t; s,\ t=1,\ldots,T.\end{aligned}$$

Because

$$\begin{aligned}E(y_{it}^2|y_{i0}) &= E((\rho^t y_{i0} + \frac{1-\rho^t}{1-\rho}a_i + \textstyle\sum_{j=1}^t, \rho^{j-1}\varepsilon_{i,t-j+1})^2|y_{i0})\\ &= \rho^{2t}y_{i0}^2 + (\frac{1-\rho^t}{1-\rho})^2(E(a_i|y_{i0})^2 + E(c_i^2|y_{i0}))+\\ &\quad E((\textstyle\sum_{j=1}^t \rho^{j-1}\varepsilon_{i,t-j+1})^2|y_{i0})\\ &= \rho^{2t}y_{i0}^2 + (\frac{1-\rho^t}{1-\rho})^2((\alpha_0+\alpha_1 y_{i0})^2 + \sigma_a^2) + \frac{1-\rho^{2t}}{1-\rho^2}(E(\varepsilon_{it}^2|y_{i0}))\\ &= \rho^{2t}y_{i0}^2 + (\frac{1-\rho^t}{1-\rho})^2((\alpha_0+\alpha_1 y_{i0})^2 + \sigma_a^2) + \frac{1-\rho^{2t}}{1-\rho^2}\sigma_\varepsilon^2,\\ E(y_{it}|y_{i0})^2 &= (\rho^t y_{i0} + \frac{1-\rho^t}{1-\rho}(\alpha_0+\alpha_1 y_{i0}))^2\\ &= \rho^{2t}y_{i0}^2 + (\frac{1-\rho^t}{1-\rho})^2((\alpha_0+\alpha_1 y_{i0})^2),\end{aligned}$$

$E(y_{it}^2|y_{i0}) - E(y_{it}|y_{i0})^2$

$$\begin{aligned}&= \rho^{2t}y_{i0}^2 + (\frac{1-\rho^t}{1-\rho})^2((\alpha_0+\alpha_1 y_{i0})^2 + \sigma_a^2) + \frac{1-\rho^{2t}}{1-\rho^2}\sigma_\varepsilon^2 -\\ &\quad \rho^{2t}y_{i0}^2 - (\frac{1-\rho^t}{1-\rho})^2((\alpha_0+\alpha_1 y_{i0})^2)\\ &= (\frac{1-\rho^t}{1-\rho})^2\sigma_a^2 \;+\; (\frac{1-\rho^{2t}}{1-\rho^2})\sigma_\varepsilon^2.\end{aligned}$$

By similar manipulation, we can express the $E(y_{it}y_{is}|y_{i0})$ - $E(y_{it}|y_{i0})E(y_{is}|y_{i0})$ in terms of parameters, ρ, σ_a^2, σ_ε^2, and time period t and s as $(\frac{1-\rho^s}{1-\rho}\frac{1-\rho^t}{1-\rho})\sigma_a^2 + \rho^{|t-s|}(\frac{1-\rho^{2s}}{1-\rho^2})\sigma_\varepsilon^2$.

If the model includes the exogenous variable x_{it} for $t = 1, \ldots, T$, the conditional mean will be

$$E(y_i|y_{i0}, x_i) = \begin{pmatrix} \vdots \\ \frac{1-\rho^t}{1-\rho}\alpha_0 + (\frac{1-\rho^t}{1-\rho}\alpha_1 + \rho)y_{i0} + \sum_{j=1}^t \rho^{j-1}x_{i,t-j+1}\beta \\ \vdots \end{pmatrix}_{T\times 1}, \text{ where}$$

$t = 1,\ldots,T$. It, however, can be easily shown that $\mathrm{Var}(y_{it}|y_{i0}, x_i)$ of (2.32) in which x_i is assumed to be strictly exogenous, is the same as that of the basic model without no other regressors beyond $y_{i,t-1}$. The illustration is in the following.

$$
\begin{aligned}
&E(y_{it}^2|y_{i0},x_i)\\
&= E((\rho^t y_{i0} + \frac{1-\rho^t}{1-\rho}a_i + \sum_{j=1}^t \rho^{j-1}(x_{i,t-j+1}\beta + \varepsilon_{i,t-j+1}))^2|y_{i0},x_i)\\
&= \rho^{2t}y_{i0}^2 + (\frac{1-\rho^t}{1-\rho})^2(E(a_i|y_{i0},x_i)^2 + E(c_i^2|y_{i0},x_i)) +\\
&\quad 2(\rho^t y_{i0} + \frac{1-\rho^t}{1-\rho}(E(a_i|y_{i0},x_i)(\sum_{j=1}^t \rho^{j-1}E(x_{i,t-j+1})\beta)) +\\
&\quad E((\sum_{j=1}^t \rho^{j-1}(x_{i,t-j+1}\beta + \varepsilon_{i,t-j+1}))^2|y_{i0},x_i)\\
&= \rho^{2t}y_{i0}^2 + (\frac{1-\rho^t}{1-\rho})^2((\alpha_0 + \alpha_1 y_{i0} + \alpha_2\overline{x}_i)^2 + \sigma_a^2) + \frac{1-\rho^{2t}}{1-\rho^2}(E(\varepsilon_{it}^2|y_{i0},x_i)) +\\
&\quad 2(\rho^t y_{i0} + \frac{1-\rho^t}{1-\rho}((\alpha_0 + \alpha_1 y_{i0} + \alpha_2\overline{x}_i)(\sum_{j=1}^t \rho^{j-1}E(x_{i,t-j+1})\beta)) +\\
&\quad (\sum_{j=1}^t \rho^{j-1}E(x_{i,t-j+1})\beta)^2\\
&= \rho^{2t}y_{i0}^2 + (\frac{1-\rho^t}{1-\rho})^2((\alpha_0 + \alpha_1 y_{i0} + \overline{x}_i)^2 + \sigma_a^2) + \frac{1-\rho^{2t}}{1-\rho^2}\sigma_\varepsilon^2 +\\
&\quad 2(\rho^t y_{i0} + \frac{1-\rho^t}{1-\rho}((\alpha_0 + \alpha_1 y_{i0} + \alpha_2\overline{x}_i)(\sum_{j=1}^t \rho^{j-1}E(x_{i,t-j+1})\beta)) +\\
&\quad (\sum_{j=1}^t \rho^{j-1}E(x_{i,t-j+1})\beta)^2.
\end{aligned}
$$

$$
\begin{aligned}
&E(y_{it}|y_{i0},x_i)^2\\
&= (\rho^t y_{i0} + \frac{1-\rho^t}{1-\rho}(\alpha_0 + \alpha_1 y_{i0} + \alpha_2\overline{x}_i) + (\sum_{j=1}^t \rho^{j-1}E(x_{i,t-j+1})\beta))^2\\
&= \rho^{2t}y_{i0}^2 + (\frac{1-\rho^t}{1-\rho})^2((\alpha_0 + \alpha_1 y_{i0} + \alpha_2\overline{x}_i)^2)+\\
&\quad 2(\rho^t y_{i0} + \frac{1-\rho^t}{1-\rho}((\alpha_0 + \alpha_1 y_{i0} + \alpha_2\overline{x}_i)(\sum_{j=1}^t \rho^{j-1}E(x_{i,t-j+1})\beta)) +\\
&\quad (\sum_{j=1}^t \rho^{j-1}E(x_{i,t-j+1})\beta)^2.
\end{aligned}
$$

Therefore, $E(y_{it}^2|y_{i0},x_i) - E(y_{it}|y_{i0},x_i)^2$ is equal to $(\frac{1-\rho^t}{1-\rho})^2\sigma_a^2 + (\frac{1-\rho^{2t}}{1-\rho^2})\sigma_\varepsilon^2$. By similar manipulation, $E(y_{it}y_{is}|y_{i0},X_{i,\max[t,s]})$ - $E(y_{it}|y_{i0},X_{it})E(y_{is}|y_{i0},X_{is})$ is easily proved to be equal to $(\frac{1-\rho^s}{1-\rho}\frac{1-\rho^t}{1-\rho})\sigma_a^2 + \rho^{|t-s|}(\frac{1-\rho^{2s}}{1-\rho^2})\sigma_\varepsilon^2$.

For model (3.1) or (3.7), the conditional mean can be derived in the way similar to the previous. Similarly, we assume that

$$a_i|y_{i0} \sim \text{Normal}(\alpha_0 + \alpha_1 \text{y}_{i0}, \sigma_{\text{a}}^2),$$

$$\varepsilon_{it}|y_i, a_i \sim \text{Normal}(0, \sigma_\varepsilon^2).$$

We can write the expectation of T observaitons on individual i conditional on y_{i0} as follows:

$E(y_i|y_{i0}) = \begin{pmatrix} \alpha_0 + (\alpha_1 + \delta_i)y_{i0} & \dots & \frac{1-\delta_i^t}{1-\delta_i}\alpha_0 + (\frac{1-\delta_i^t}{1-\delta_i}\alpha_1 + \delta_i^t)y_{i0} & \dots \end{pmatrix}_{T\times 1}$, where $t = 1, \dots, T$ and $\delta_i \equiv \rho + \gamma(\alpha_0 + \alpha_1 y_{i0} + c_i)$.

$$\begin{aligned} \Omega(y_{i0}) &\equiv Var(y_i|y_{i0}) \\ &= E\left((y_i - E(y_i|y_{i0}))(y_i - E(y_i|y_{i0}))'|y_{i0}\right) \\ &= E(y_i y_i'|y_{i0}) - E(y_i|y_{i0})E(y_i|y_{i0})' \end{aligned}$$

$$\Omega(y_{i0}) = \begin{pmatrix} \omega_{11} & \dots & \omega_{1T} \\ \vdots & \ddots & \vdots \\ \omega_{T1} & \dots & \omega_{TT} \end{pmatrix},$$

where $\omega_{tt} = E(y_{it}^2|y_{i0}) - E(y_{it}|y_{i0})^2$ and $\omega_{st} = E(y_{is}y_{it}|y_{i0}) - E(y_{is}|y_{i0})E(y_{it}|y_{i0})$. The elements of $\Omega(y_{i0})$ are as follows:

$$\omega_{tt} = (\frac{1-\delta_i^t}{1-\delta_i})^2\sigma_a^2 + (\frac{1-\delta_i^{2t}}{1-\delta_i^2})\sigma_\varepsilon^2,\ t = 1, \dots, T,$$

$$\omega_{st} = (\frac{1-\delta_i^s}{1-\delta_i}\frac{1-\delta_i^t}{1-\delta_i})\sigma_a^2 + \delta_i^{|t-s|}(\frac{1-\delta_i^{2s}}{1-\delta_i^2})\sigma_\varepsilon^2,\ s \neq t; s,\ t = 1, \dots, T.$$

Because

$$\begin{aligned} E(y_{it}^2|y_{i0}) &= E((\delta_i^t\, y_{i0} + \frac{1-\delta_i^t}{1-\delta_i}a_i + \textstyle\sum_{j=1}^t \delta_i^{j-1}\, \varepsilon_{i,t-j+1})^2|y_{i0}) \\ &= \delta_i^{2t}y_{i0}^2 + (\frac{1-\delta_i^t}{1-\delta_i})^2(E(a_i|y_{i0})^2 + E(c_i^2|y_{i0})) + \\ &\quad E((\textstyle\sum_{j=1}^t \delta_i^{j-1}\, \varepsilon_{i,t-j+1})^2|y_{i0}) \\ &= \delta_i^{2t}y_{i0}^2 + (\frac{1-\delta_i^t}{1-\delta_i})^2((\alpha_0 + \alpha_1 y_{i0})^2 + \sigma_a^2) + \frac{1-\delta_i^{2t}}{1-\delta_i^2}(E(\varepsilon_{it}^2|y_{i0})) \\ &= \delta_i^{2t}y_{i0}^2 + (\frac{1-\delta_i^t}{1-\delta_i})^2((\alpha_0 + \alpha_1 y_{i0})^2 + \sigma_a^2) + \frac{1-\delta_i^{2t}}{1-\delta_i^2}\sigma_\varepsilon^2, \\ E(y_{it}|y_{i0})^2 &= (\delta_i^t y_{i0} + \frac{1-\delta_i^t}{1-\delta_i}(\alpha_0 + \alpha_1 y_{i0}))^2 \\ &= \delta_i^{2t}y_{i0}^2 + (\frac{1-\delta_i^t}{1-\delta_i})^2((\alpha_0 + \alpha_1 y_{i0})^2), \end{aligned}$$

$E(y_{it}^2|y_{i0}) - E(y_{it}|y_{i0})^2$

$$
\begin{aligned}
&= \delta_i^{2t} y_{i0}^2 + (\frac{1-\delta_i^t}{1-\delta_i})^2((\alpha_0 + \alpha_1 y_{i0})^2 + \sigma_a^2) + \frac{1-\delta_i^{2t}}{1-\delta_i^2}\sigma_\varepsilon^2 - \\
&\quad \delta_i^{2t} y_{i0}^2 - (\frac{1-\delta_i^t}{1-\delta_i})^2((\alpha_0 + \alpha_1 y_{i0})^2) \\
&= (\frac{1-\delta_i^t}{1-\delta_i})^2 \sigma_a^2 \; + \; (\frac{1-\delta_i^{2t}}{1-\delta_i^2})\sigma_\varepsilon^2.
\end{aligned}
$$

We can express the $E(y_{it}y_{is}|y_{i0})$ - $E(y_{it}|y_{i0})E(y_{is}|y_{i0})$ in terms of parameters, ρ, σ_a^2, σ_ε^2, and time period t and s as $(\frac{1-\delta_i^s}{1-\delta_i}\frac{1-\delta_i^t}{1-\delta_i})\sigma_a^2 + \delta_i^{|t-s|}(\frac{1-\delta_i^{2s}}{1-\delta_i^2})\sigma_\varepsilon^2$.

If the model includes the exogenous variable x_{it} for $t = 1, \ldots, T$, we replace the assumptions on a_i and ε_{it} with

$$
\begin{aligned}
&a_i|y_{i0}, x_i \sim \text{Normal}(\alpha_0 + \alpha_1 \mathrm{y_{i0}} + \overline{\mathrm{x}}_\mathrm{i}\, \alpha_2, \sigma_\mathrm{a}^2), \\
&\varepsilon_{it}|y_i, a_i \sim \text{Normal}(0, \sigma_\varepsilon^2).
\end{aligned}
$$

The conditional mean will be

$$
E(y_i|y_{i0}, x_i) = \begin{pmatrix} \vdots \\ \frac{1-\delta_i^t}{1-\delta_i}\alpha_0 + (\frac{1-\delta_i^t}{1-\delta_i}\alpha_1 + \delta_i^t) y_{i0} + \sum_{j=1}^t \delta_i^{j-1} x_{i,t-j+1}\beta \\ \vdots \end{pmatrix}_{T\times 1},
$$

where $\delta_i \equiv \rho + \gamma(\alpha_0 + \alpha_1 y_{i0} + \overline{x}_i + c_i)$ and $t = 1, \ldots, T$.

$\mathrm{Var}(y_{it}|y_{i0}, x_i)$ is different from that of the basic model without no other regressors beyond $y_{i,t-1}$. The illustration is in the following.

$$
\begin{aligned}
&E(y_{it}^2|y_{i0}, x_i) \\
&= E((\delta_i^t y_{i0} + \frac{1-\delta_i^t}{1-\delta_i} a_i + \textstyle\sum_{j=1}^t \delta_i^{j-1}(x_{i,t-j+1}\beta + \varepsilon_{i,t-j+1}))^2|y_{i0}, x_i) \\
&= \delta_i^{2t} y_{i0}^2 + (\frac{1-\delta_i^t}{1-\delta_i})^2 (E(a_i|y_{i0}, x_i)^2 + E(c_i^2|y_{i0}, x_i)) + \\
&\quad 2(\delta_i^t y_{i0} + \frac{1-\delta_i^t}{1-\delta_i}(E(a_i|y_{i0}, x_i)(\textstyle\sum_{j=1}^t \delta_i^{j-1} E(x_{i,t-j+1})\beta)) + \\
&\quad E((\textstyle\sum_{j=1}^t \delta_i^{j-1}(x_{i,t-j+1}\beta + \varepsilon_{i,t-j+1}))^2|y_{i0}, x_i) \\
&= \delta_i^{2t} y_{i0}^2 + (\frac{1-\delta_i^t}{1-\delta_i})^2((\alpha_0 + \alpha_1 y_{i0} + \alpha_2 \overline{x}_i)^2 + \sigma_a^2) + \frac{1-\delta_i^{2t}}{1-\delta_i^2}(E(\varepsilon_{it}^2|y_{i0}, x_i)) +
\end{aligned}
$$

$$2(\delta_i^t y_{i0} + \frac{1-\delta_i^t}{1-\delta_i}((\alpha_0 + \alpha_1 y_{i0} + \alpha_2 \overline{x}_i)(\textstyle\sum_{j=1}^t \delta_i^{j-1} E(x_{i,t-j+1})\beta)) +$$

$$(\textstyle\sum_{j=1}^t \delta_i^{j-1} E(x_{i,t-j+1})\beta)^2$$

$$= \delta_i^{2t} y_{i0}^2 + (\frac{1-\delta_i^t}{1-\delta_i})^2((\alpha_0 + \alpha_1 y_{i0} + \overline{x}_i)^2 + \sigma_a^2) + \frac{1-\delta_i^{2t}}{1-\delta_i^2}\sigma_\varepsilon^2 +$$

$$2(\delta_i^t y_{i0} + \frac{1-\delta_i^t}{1-\delta_i}((\alpha_0 + \alpha_1 y_{i0} + \alpha_2 \overline{x}_i)(\textstyle\sum_{j=1}^t \delta_i^{j-1} E(x_{i,t-j+1})\beta)) +$$

$$(\textstyle\sum_{j=1}^t \delta_i^{j-1} E(x_{i,t-j+1})\beta)^2.$$

$$E(y_{it}|y_{i0}, x_i)^2$$

$$= (\delta_i^t y_{i0} + \frac{1-\delta_i^t}{1-\delta_i}(\alpha_0 + \alpha_1 y_{i0} + \alpha_2 \overline{x}_i) + (\textstyle\sum_{j=1}^t \delta_i^{j-1} E(x_{i,t-j+1})\beta))^2$$

$$= \delta_i^{2t} y_{i0}^2 + (\frac{1-\delta_i^t}{1-\delta_i})^2((\alpha_0 + \alpha_1 y_{i0} + \alpha_2 \overline{x}_i)^2) +$$

$$2(\delta_i^t y_{i0} + \frac{1-\delta_i^t}{1-\delta_i}((\alpha_0 + \alpha_1 y_{i0} + \alpha_2 \overline{x}_i)(\textstyle\sum_{j=1}^t \delta_i^{j-1} E(x_{i,t-j+1})\beta)) +$$

$$(\textstyle\sum_{j=1}^t \delta_i^{j-1} E(x_{i,t-j+1})\beta)^2.$$

Therefore, $E(y_{it}^2|y_{i0}, x_i) - E(y_{it}|y_{i0}, x_i)^2$ is equal to $(\frac{1-\delta_i^t}{1-\delta_i})^2\sigma_a^2 + (\frac{1-\delta_i^{2t}}{1-\delta_i^2})\sigma_\varepsilon^2$. By similar manipulation, $E(y_{it}y_{is}|y_{i0}, X_{i,\max[t,s]})$ - $E(y_{it}|y_{i0}, x_i)E(y_{is}|y_{i0}, X_{is})$ is easily proved to be equal to $(\frac{1-\delta_i^s}{1-\delta_i}\frac{1-\delta_i^t}{1-\delta_i})\sigma_a^2 + \delta_i^{|t-s|}(\frac{1-\delta_i^{2s}}{1-\delta_i^2})\sigma_\varepsilon^2$.

Therefore, when the state dependence interacts with the unobserved effect, the autoregressive coefficient contains the unobserved effect and hence the conditional mean and variance depend on y_{i0} and $\overline{x}_i$.

Appendix B

IV estimator for average autoregressive coefficient across population of unobserved heterogeneity

In this appendix we prove that the IV estimator for dynamic model where the state dependence depends on the unobserved effects is not consistent We define some notations for polynomial in the process of proof for the simplicity of exposition as follows:

1. Let $\mathbf{b} \in \mathbb{R}^{n+1}$ be a non-zero coefficient vector $\mathbf{b} := (b_0, b_1, \ldots, b_n)$. Denote $\mathbf{P_n}$ a non-trivial polynomial of degree n :

$$\mathbf{P_n}(x, \mathbf{b}) \equiv b_0 + b_1 x + b_2 x^2 + \cdots + b_n x^n \in \mathbb{R}[x].$$

2. For $m \in \mathbb{R}$, $\lfloor m \rfloor := \{n \mid \text{if } n \leq m < n+1 \text{ for some } n \in \mathbb{Z}\}$.

According to the first difference equation of (3.1), we use the lagged $\Delta y_{i,t-2}$ as an instrument for $\Delta\, y_{i,t-1}$ and we can write the equation as follows:

$$E(\Delta y_{it} y_{i,t-2}) = E(\vartheta_i\, \Delta y_{i,t-1}\, \Delta y_{i,t-2}) \;+\; E(\Delta\varepsilon_{it}\, \Delta y_{i,t-2}), \tag{B.1}$$

where $\vartheta_i = \rho + \gamma\, a_i$. We make the following assumptions:

Assumption B.1 : $E(\varepsilon_{it} | y_{i,t-1}, \ldots, y_{i0}, a_i) = 0.$

Assumption B.2 : $Var(\varepsilon_{it} | y_{i,t-1}, \ldots, y_{i0}) \;=\; \sigma^2_{\varepsilon}.$

Assumption B.3 : $E(a_i | y_{i,t-1}, \ldots, y_{i0}) = E(a_i).$

Assumption B.4 : $Var(a_i | y_{i,t-1}, \ldots, y_{i0}) \;=\; \sigma^2_{a}.$

Assumption B.5 : $E(y_{i0}) = \mu_{y0}$ and $Var(y_{i0}) = \sigma^2_{y0}.$

According to Assumption B.1 and B.3, equation (B.1) is equal to

$$E(\Delta y_{it} y_{i,t-2}) = \vartheta\, E(\Delta y_{i,t-1}\, \Delta y_{i,t-2}) \;+ \gamma\, E(c_i\, \Delta y_{i,t-1} y_{i,t-2}), \tag{B.2}$$

where $\vartheta = \rho + \gamma\, \mu_a$. The IV estiamtor is as follows:

$$\begin{aligned} \hat{\vartheta}_{IV} &= \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \Delta y_{i,t} \Delta y_{i,t-2}}{\sum_{i=1}^{N} \sum_{t=1}^{T} \Delta y_{i,t-1} \Delta y_{i,t-2}} \\ &= \vartheta + \frac{\gamma \sum_{i=1}^{N} \sum_{t=1}^{T} c_i \Delta y_{i,t-1} \Delta y_{i,t-2}}{\sum_{i=1}^{N} \sum_{t=1}^{T} \Delta y_{i,t-1} \Delta y_{i,t-2}} + \\ & \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \Delta\varepsilon_{it} \Delta y_{i,t-2}}{\sum_{i=1}^{N} \sum_{t=1}^{T} \Delta y_{i,t-1} \Delta y_{i,t-2}}. \end{aligned} \tag{B.3}$$

where $\vartheta \;= \rho \;+\; \gamma\, \mu_a$. Under Assumption 1 to 5, the probability limit of ϑ_{IV} is as follows:

$$\begin{aligned} \operatorname*{plim}_{N\to\infty} \hat{\vartheta}_{IV} &= \frac{\operatorname*{plim}_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \Delta y_{i,t} \Delta y_{i,t-2}}{\operatorname*{plim}_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \Delta y_{i,t-1} \Delta y_{i,t-2}} \\ &= \vartheta + \frac{\gamma \operatorname*{plim}_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} c_i \Delta y_{i,t-1} \Delta y_{i,t-2}}{\operatorname*{plim}_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \Delta y_{i,t-1} \Delta y_{i,t-2}} + \\ & \frac{\operatorname*{plim}_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \Delta\varepsilon_{it} \Delta y_{i,t-2}}{\operatorname*{plim}_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \Delta y_{i,t-1} \Delta y_{i,t-2}}. \end{aligned} \tag{B.4}$$

The proof of $\operatorname*{plim}_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} c_i \Delta y_{i,t-1} \Delta y_{i,t-2} \neq 0$ is in the fowlling.

$$\operatorname*{plim}_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} c_i \Delta y_{i,t-1} \Delta y_{i,t-2} = \sum_{t=3}^{T} E\left(c_i \Delta y_{i,t-1} \Delta y_{i,t-2}\right)$$
$$= \sum_{t=3}^{T} \left(\underbrace{E(c_i y_{i,t-1} y_{i,t-2})}_{(A1)} - \underbrace{E(c_i \Delta y_{i,t-3}^2)}_{(A2)} \right). \tag{B.5}$$

According to Assumption B.1 to B.5, equation (B.5) is solved out as follows:

$$(A1) = \sum_{t=3}^{T} \left(E\left(c_i \vartheta y_{i,t-2}^2 + c_i \gamma y_{i,t-2}^2 + c_i (\mu_a + c_i) y_{i,t-2} + c_i \varepsilon_{i,t-1} y_{i,t-2}\right)\right)$$
$$= \sum_{t=3}^{T} E(c_i \vartheta y_{i,t-2}^2) + \sum_{t=3}^{T} E(c_i \gamma y_{i,t-2}^2) + \sum_{t=3}^{T} E(c_i (\mu_a + c_i) y_{i,t-2}) \tag{B.6}$$

Putting the final expression of (B.6) into (A1) in (B.5), we can obtain the following equation:

$$\sum_{t=3}^{T} E(c_i \Delta y_{i,t-1} \Delta y_{i,t-2}) = (\vartheta - 1 - \gamma) \sum_{t=3}^{T} E\left(c_i y_{i,t-2}^2\right) + \sum_{t=3}^{T} E\left(c_i \gamma y_{i,t-2}^2\right)$$
$$+ \sum_{t=3}^{T} E\left(c_i (\mu_a + c_i) y_{i,t-2}\right) \tag{B.7}$$

We can seperately solve out the three terms of (B.7), $\sum_{t=3}^{T} E\left(c_i y_{i,t-2}^2\right)$, $\sum_{t=3}^{T} E\left(c_i \gamma y_{i,t-2}^2\right)$ as well as $\sum_{t=3}^{T} E\left(c_i (\mu_a + c_i) y_{i,t-2}\right)$.

$$\sum_{t=3}^{T} E\left[c_i y_{i,t-2}^2\right] = \sum_{t=3}^{T} E\left[c_i \left(\vartheta_i^{t-2} y_{i0} + \frac{1-\theta_i^{t-2}}{1-\theta_i} a_i + \sum_{j=1}^{t-2} \vartheta_i^{j-1} \varepsilon_{i,t-j-1} \right)^2 \right]$$
$$= \sum_{t=3}^{T} \sigma_a^2 (\sigma_{y0}^2 + \mu_{y0}^2) E(y_{i0}^2) \mathbf{P}_{t-2}(\sigma_a^2, \mathbf{b}) + \sum_{t=3}^{T} \sigma_a^2 \mathbf{P}_{\mathbf{t-2}}(\sigma_a^2, \mathbf{d})$$
$$+ 2 \sum_{t=3}^{T} \mu_{y0} \sigma_a^2 \mathbf{P}_{\mathbf{t-2}}(\sigma_a^2, \mathbf{e}). \tag{B.8}$$

By the manipulation similar to (B.8), we can express the other terms, $\sum_{t=3}^{T} E\left(c_i \gamma y_{i,t-2}^2\right)$ in terms of (t-2)-order polynomial of σ_c^2 and $\sum_{t=3}^{T} E\left(c_i (\mu_a + c_i) y_{i,t-2}\right)$ in terms of ($\lfloor \frac{t-1}{2} \rfloor$)-order polynomial of σ_c^2, and their relevant coefficients are function $(\rho, \mu_a, \mu_{y0}, \sigma_a^2, \sigma_\varepsilon^2, \sigma_{y0}^2)$. Therefore, $\operatorname*{plim}_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} c_i \Delta y_{i,t} y_{i,t-2}$ is not zero when T is fixed.

Similarly, we can prove that

$$(1) \quad \operatorname*{plim}_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \Delta\varepsilon_{it} y_{i,t-2} = 0$$

$$(2) \quad \operatorname*{plim}_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \Delta y_{i,t-1} y_{i,t-2} \neq 0.$$

As we have shown, Assumption B.1 to B.5 can not lead to the fact that $\hat{\vartheta}_{IV}$ is consistent. Even more,if we make the extreme assumption that c_i, ε_{it} and y_{i0} are independent with each other, then $\operatorname*{plim}_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} c_i \Delta y_{i,t-1} \Delta y_{i,t-2}$ is not equal to zero. It is obvious that such a strong assumption appears that many expectation of terms in the right hand side of equation of (B.7) vanished because any term in the form of $E[\mathbf{P}_s((c_i^{m_1} \varepsilon_{it}^{m_2} y_{i0}^{m_3}))]$ is zero; there, nevertheless, exist such non-zero terms $E(\mathbf{P}_m(c_i)c_i^k)$, where k is 0 or 1, since $m > 2$ and hence a standard IV method applied to first-differenced equation can not consistently estimate the parameter of interest.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] **Amemiya Y. (1985)**. Instrumental variable estimator for the nonlinear error-in-variables model. Journal of Econometrica, 28, 273-290.

[2] **Amemiya Y (1990)**. Two-stage instrmental variable estimators for the nonlinear-error-in-variables model. Journal of Econometrics, 44, 273-290.

[3] **Abramowitz, M. and I. Stegun (1964)**. Handbook of mathematical Functions with formulas, graphs and mathematical tables. National Bureau of Standards Applied Mathematics Series. No. 55. Washington, D.C.: US. Government Printing Office.

[4] **Aigner, D.J., Hsiao Cheng, A. Kapteyn and T. Wansbeek (1984)**. Latent variable midels in econometrics, in Handbook of Econometrics, vol.2, edited by Z. Griliches and M. Intriligator, 1322-1393. Amsterdam: North-Holland.

[5] **Anderson T.W. and Hsiao Cheng (1982)**. Formulation and estimation of dynamic models using panel data. Journal of Econometrics, 18, 47 - 82.

[6] **Anderson, T.W. and Hsiao Cheng, (1981a)**. Estimation of dynamic models with error components. Journal of the American Statistical Association, 76, 598-606.

[7] **Arellano, M. (1988)**. A note on the Anderson-Hsiao estimator for panel data; mimeo Institute of Economics, Oxford University.

[8] **Arellano, M. and Bond, S.R., (1991)**. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies, 58, 277-297.

[9] **Balestra, P., and M. Nerlove (1966)**. Pooling cross-sectin and time-series data in the estimation of a dynamic model. Econometrica, 34, 585-612.

[10] **Bhargava, A., and J.D Sargan (1983)**. Estimating Dynamic random effects models from Panel data covering short time periods. Econometrica 51, 1635-59.

[11] **Bollerslev Tim, Wooldridge J.M. (1992)**. Quasi-maximum likelihood estimation and inference in dynamic models wint time-varying covariances. Econometric Reviews, 11(2), 143-172.

[12] **Blundell Richard and Bond Stephon (1998)**. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics, 87, 115-143.

[13] **Berndt E.K., Hall B.H., Hall R.E. and Hausman J.A. (1974)**. Estimation and inefficient in nonlinear structual models. Annals of Econometric and Social Measurement, 3/4, 653-665.

[14] **Butter, J.S. and Bobert Moffitt (1982)**. A computationally efficient quarture procedure for the one-factor multinomial probit model. Econometric, 50,761-765.

[15] **Davidson, R. and J.G. Mackinnin (1984)**. Convenient specification tests for Logit and Probit models. Journal of Econometrics, 34, 241-262.

[16] **Gallant, A.R. and D.W. Nychka (1987)**. Semi-nonparametric Maximum Likelihood Estimation. Econometrica, 55, 363-390.

[17] **Gary Chamberlain (1987)** Asymptotic efficiency in estimation with conditinal moment restrictions, Journal of Econometrics, 34, 305-334.

[18] **Gary Chamberlain (1992)**. Comment: Sequential moment restrictions in panel data, Journal of Business and Economic Statistics., 10, 20-26.

[19] **Gourieroux, C.A. Monfort and A. Trognon (1984)**. Pseudo-Maximum Likelihood methods: Theory. Econometrics, 52, 681-700.

[20] **Hahn Jinyong (1999)**. How informative is the initial condition in the dynamic panel model with fixed effects. Journal of Econometrics, 93, 309-326.

[21] **Hammer, P.C. and Stroud, A.J. (1956)** Numerical integration over simplexes, Math. Tables Aids Comp.,10, 137-139.

[22] **Hausman, J. A. and W. E. Taylor (1981)**. Panel data and unobservable individual effects. Econometrica, 49, 1377-1338.

[23] **Holtz-Eakin, D., W.K. Newey and H.S. Rosen (1988)**. Estimating vector autoregressions with panel data. Econometrica, 56, 1371-1395.

[24] **Honorè, B.E. (1993)**. Orthogonality Conditions for Tobit Models wiht Fixed Effects and Lagged Dependent Variables. Journal of Econometricas 59, 35-61.

[25] **Honorè, B.E. and E. Kyriadizou (2000).** Panel Data Discrete Choice Models with Lgged Dependent Variables. Econometrica, Vol.68, No.4 839 - 874.

[26] **Hsiao Cheng (1986).** Analysis of panel data, Cambridge: Cambridge University press.

[27] **Javier Alvarez and Mannuel Arellano (1998).** The time series and cross-section asymptotics of dynamic panel data eseimators. Working paper, No9808.

[28] **Keane, M.P. (1993)** Simulation estimation for panel data models with limited dependent variables. Handbook of Statistics, Volume 11, G.s. Maddala, C.R. Rao, H.D. Vinod (eds.) Amsterdam: North Holland, 545-571.

[29] **Kyung So Im, Seung, C. Ahn, Schmidt, P. and J.M. Wooldridge(1999).** Efficient estimation of panel data models with strictly exogenous explanatory variables. Journal of Econometrics, 93, 177-201.

[30] **Laszio Matyas and Patrick Sevestre (1992).** The Econometrics of panel data, handbook of theory and applications. Kluwer Academic publishers.

[31] **Maddala, G. S. (1971).** The use of variance components models in pooling cross-section and time series data. Econometrica,39, 341-358.

[32] **Manski, C.F. (1987).** Semiparametric analysis of random effects linear models from binary panel data. Econometrica,55, 357-362.

[33] **Manski, C.F. (1988).** Analog estimation methods in econometrics, New York: Chapman and Hall.

[34] **Nerlove, M., (1971).** Further evidence on the estimation o f dynamic relations from a time series of cross sections. Econometric, 39, 359-382.

[35] **Neyman, J., and E. L. Scott (1948).** Consistent estimates based on partially consistent observations. Econometrica,16, 1-32.

[36] **Newey, W.K. and D. McFadden (1994).** Large sample estimation and hypothesis testing. in R.F. Engle and D. McFadden (eds.), Handbook of Econometrics, Volume 4. Amsterdam: North Holland, 2111-2245.

[37] **Nickell, S. (1981).** Biases in dynamic models with fixed effects, Econometrica, 49, 1417-1426.

[38] **Patric, Sevestre and Alain, Trognon (1992).** Linear dynamic models, in László Mátyás, Patric Sevestre (eds.), The Econometrics of Panel Data, Handbook of Theory and Applications, 95-117.

[39] **Stroud, A.H. and Secrest, Don (1966).** Gaussian quadrature formulas, Englewood Cliffs: Prentice-Hal.

[40] **Seung, C. Ahn and Schmidt Peter (1995).** Efficient estimatin fo models for dynaic panel data. Jourmnal of Econometrics, 68, 5-27.

[41] **Seung, C. Ahn and Schmidt, Peter (1997).** Efficient estimation of dynamic panel data models: Alternative assumptions and simplified estimation. Journal of Econometrics, 76, 302-321.

[42] **Trognon, A.(1978).** Miscellaneous asymptotic properties of ordinary least squares and maximum likelihood estimations in dynamic error components models. Annales de I'INSSEE, no. 30-31, 632-657. component model. International Economic Review,12, 1-13.

[43] **Vella, F. and Verbeek, M. (1988).** Whose ways do union raise? a dynamic model of unionism and wage rate determiniation for young men. Journal of Aplied Econometrics, 13, 163-188.

[44] **Wooldridge, J.M. (1994)** . Estimation and inference for dependent processes, in R.F. Engle and D. McFadden (eds.) Handbook of Econometrics, Volume IV Amsterdam: North Holland, 2639-2738.

[45] **Wooldridge, J.M. (1996).** Estimating systems of equations with different instruments for different equations. Journal of Econometrics 74, 387 - 405.

[46] **Wooldridge, J.M. (1997).** Multiplicative Panel Data models without the strict exogeneity assumption. Econometric Theory, 13, 667-678.

[47] **Wooldridge, J.M. (1999)** . Distribution-free estimation of some nonlinear panel data models. Journal of Econometrics, 90, 77-97.

[48] **Wooldridge, J.M. (1999).** Introductory econometrics: a modern approach. South Western College.

[49] **Wooldridge, J.M. (2000a).** A framework for estimating dynamic, unobserved effects panel data models with possible feedback to future explanatory variables. Econometrics Letters 68, 245-250.

[50] **Wooldridge, J.M. (2000b)**. The initial conditions problem in dynamic, non-linear Panel Data models with unobserved heterogeneity, mimeo. Michigan State University Department of Economics.

[51] **Wooldridge, J.M. (2001)**. Econometric analysis of cross section and Panel Data. forthcoming, MIT Press. Cambridge, MA.