DYNAMIC LABOR-PARTICIPATION BEHAVIOR OF KOREAN WOMEN

(May 15, 2003)

Myoung-jae Lee
School of Economics and Social Sciences
Singapore Management University
469 Bukit Timah Road
Singapore 259756
Fax: 65-6822-0833; E-mail: mjlee@smu.edu.sg

Yoon-hee Tae
School of Economics
Sungkyunkwan University
Seoul, Korea

We analyze dynamic labor participation behavior of Korean women for 1998-2001. State dependence under unobserved heterogeneity is considered, and we allow for the unobserved heterogeneity to be unrelated, pseudo-related, or related arbitrarily to regressors, which is done by dynamic probits and a recently developed dynamic conditional logit. In terms of methodological contribution, a simple three-stage algorithm is proposed for dynamic probits, which reduces the computation time about by half; it is also shown that the often-used practice of treating the initial response as fixed should not be used. The following main empirical conclusions emerged. First, the state dependence is about $0.6 \times SD(error)$, higher for married or junior-college-educated, and lower for women in 20's and 30's; state dependence is almost zero for single young women; the degree of state dependence is lower than that for developed countries. Second, while education increases participation, college education has negative effects for women in 40's or above. Third, marriage has a very high negative short-term effect but a positive long-term effect. These findings have a number of policy implications. First, rather than college education, junior college education more geared for job skills should be supported. Second, policy efforts to increase female labor participation should be directed at women of relatively high age (40 or above) or married.

JEL Classification Numbers: C33, C35, J21.

Key Words: labor participation, state dependence, panel data, conditional logit.

1. Introduction

For a choice variable, state dependence has an important implication: once a choice is made, then the subject is "hooked on" and the same choice is likely to be made in future. For female labor supply, this means, once a woman makes the choice of working (versus non-working), the simple fact that she works now will increase the likelihood that she works in future, with the other things held constant. In this case, if the government wants to increase female labor supply, then the policy effort should be directed to increasing early participation rates rather than to increasing the work hours of working females, because once women start to work they will just keep working. This has much in common with advertising effect on sales: under state dependence, consumers get hooked on the product once they consume it, and the effect of one-shot advertising will last long with lesser need for follow-up advertising.

Formally, let $1[A] = 1$ if A holds and 0 otherwise, and suppose

$$\begin{aligned} y_{it}^{*} &= \beta_y y_{i,t-1} + x_{it}'\beta + v_{it}, \;\; v_{it} = \delta_i + u_{it}, \quad y_{it} = 1[y_{it}^{*} > 0], \\ (x_{it}', y_{it}) \text{ is observed, } i &= 1, ..., N, \;\; t = 1, ..., T, \text{ iid across } i, \end{aligned} \qquad (1.1)$$

where $y_{it} = 1$ denotes that female i works at time t, x_{it} is a regressor vector with its first component being 1, δ_i is a time-constant error (also called 'individual effect' or 'unobserved heterogeneity'), u_{it} is a time-variant error, and β_y and β are conformable parameters. If $\beta_y \neq 0$, there exists a state dependence.

State dependence in female labor supply can arise for a number of reasons. First, preference for work (versus leisure) can be intertemporally non-separable: leisures at two time points can be complements or substitutes. Second, job-search cost may differ depending on work or no-work status: for working women, the search cost for the current job would be smaller than for non-working women, but the search cost for outside jobs may be higher. Third, human capital accumulation may differ: working women may have higher human capital that is good for any job, or they may accumulate only the job-specific human capital and lose the human capital good for other jobs. Fourth, the work status may have signaling (or scarring) effect about productivity of the female. Taken together, state dependence can be positive or negative. But the literature all points to positive effects of substantial magnitudes. Just like most things in life, the temporal behavior seems to be positively correlated.

Whatever the reason for state dependence may be, observationally, state dependence shows up as persistence of one choice. But there is another source of choice persistence : temporal dependence of error terms. A high $COR(v_{i,t-1}, v_{it})$ implies choice persistence, and $COR(v_{i,t-1}, v_{it})$ can be high for two reasons. One is because $v_{i,t-1}$ and v_{it} share δ_i, and the other is because $COR(u_{i,t-1}, u_{it})$ is high. Which is the more important depends on $V(\delta_i)/V(u_{it})$ (or $V(\delta_i)/\{V(\delta_i) + V(u_{it})\}$). In total, we can think of three sources for choice persistence: $\beta_y \neq 0$, $V(\delta_i)/V(u_{it}) \neq 0$, and $COR(u_{i,t-1}, u_{it}) \neq 0$; choice persistence can result also from persistence in x_{it}, but this will be ignored because x_{it} can be controlled for. For reasons to be given later, we will consider only the first two sources for our empirical work — state dependence and unobserved heterogeneity— assuming that $u_{i1}, ..., u_{iT}$ are independent.

For policy-wise, unobserved heterogeneity has a different implication from state dependence: the policy does not work unless it changes the individual heterogeneity, which would require more long-term policy outlay instead of one-shot. Separating state dependence from unobserved heterogeneity cannot be done with cross-section data nor with time-series data: it requires panel data. In the literature of panel data, there are two types of assumptions for δ_i: one is 'fixed effect' where δ_i is related to $x_{i1}, ..., x_{iT}$ arbitrarily, and the other is 'random effect' where δ_i is independent of $x_{i1}, ..., x_{iT}$. The terms 'fixed effect' and 'random effect' are, however, misnomers because they do not denote what they really are; we will use the terms 'related effect' and 'unrelated effect', respectively, following Lee (2002).

There are many papers dealing with state dependence for labor supply using panel data. We will briefly describe some recent papers in the following (more references can be found in the papers). But before we proceed, two issues deserve to be mentioned. One issue is the assumption on δ_i: in addition to related and unrelated effects, another approach suggested in Chamberlain (1984) is modeling δ_i as a linear function of $x_{i1}, ..., x_{iT}$ plus an unrelated effect η_i; this way, δ_i is allowed to be related to the regressors (but the relation is spelled out), and we will call the approach 'pseudo-related effect'. The other issue is the initial value problem of how to model the y_{i1} equation. For this, we list three approaches: for some parameters $\alpha_1, ..., \alpha_T, \alpha_\delta$, the three approaches are

(i) y_{i1} is not random to treat y_{i1} only as a regressor in the y_{i2} equation.

$$(ii)\ y_{i1} = 1[x'_{i1}\alpha_1 +, ..., +x'_{iT}\alpha_T + v_{i1} > 0], \quad COR(v_{i1}, v_{it}) \equiv \rho_v\ \forall t = 2, ..., T. \qquad (1.2)$$

$$(iii)\ y_{i1} = 1[x'_{i1}\alpha_1 +, ..., +x'_{iT}\alpha_T + \alpha_\delta\delta_i + u_{i1} > 0], \quad COR(u_{i1}, u_{it}) = 0\ \forall t = 2, ..., T.$$

The first is the simplest but unrealistic. The second is general, but difficult to implement requiring a high-dimensional integration. The third that falls in between (i) and (ii) in terms of its strength of assumptions is frequently used in practice as to be seen shortly; $V(u_{i1}) \equiv \sigma_1^2$ in (iii) is allowed to differ from $V(u_{it}) \equiv \sigma_u^2 \ \forall t = 2, ..., T$. In (ii) and (iii), $\alpha_1, ..., \alpha_T$ may get further restricted in practice; for instance, $\alpha_t = 0 \ \forall t \neq 1$. For pseudo-related effects, δ_i in (iii) is replaced by η_i. Normality assumption has been used for the error terms; for this reason, we will call a model with (1.1) and (1.2) a "dynamic probit" from now on.

Shaw (1994) uses PSID over 1967-1987 for the U.S. white females; Shaw use three separate data (single, single-to-married, and married) and allow different coefficients for four different age groups. Shaw estimates labor participation and work hour equations separately with lagged work-hour, not lagged participation, in both equations; because of this aspect, this study is not quite relevant to our study.

Mühleisen and Zimmermann (1994) use the German Socio-Economic Panel over 1984-1989 for German males. They use an unrelated-effect Maximum Likelihood Estimator (MLE) under normality assumptions, allowing for $COR(u_{i,t-1}, u_{it}) \neq 0$ with $u_{it} = \xi \cdot u_{i,t-1} + \varepsilon_{it}$ where ξ is a parameter and ε_{it}'s are iid across i and t. This generality, however, poses a multidimensional integration problem in getting the likelihood function, for which a method of simulated likelihood is used. Mühleisen and Zimmermann's response variable is $1 - y_{i,t}$ and they let $1 - y_{i,t-1}$ interact with 6 regressors; the estimate for $1 - y_{i,t-1}$ is 3.31 (significant). They find ξ to be about -0.32 (significant). But their estimate for $V(\delta_i)/V(\varepsilon_{it}) = 0.023$ (insignificant), which means $V(\delta_i)/V(u_{it}) = 0.023/(1 - (-0.32)^2) \simeq 0.026$; that is the unobserved heterogeneity is almost nonexistent. In a strict sense, this means degeneracy in their MLE.

Hyslop (1999) uses PSID over 1979-1985 for married females. Hyslop uses a pseudo-related effect MLE where δ_i is specified as a linear function of elements of $x_{i1}, ..., x_{iT}$ plus an unrelated-effect η_i. The linear function is then absorbed into the regression function of y^*_{it}, and $V(\eta_i)/\{V(\eta_i) + V(u_{it})\}$ is identified. The initial value problem is deal with as in (ii) above, allowing for $COR(u_{i,t-1}, u_{it}) \neq 0$ with $u_{it} = \xi \cdot u_{i,t-1} + \varepsilon_{it}$. The multidimensional integration problem due to (ii) and $u_{it} = \xi u_{i,t-1} + \varepsilon_{it}$ is handled with a simulated likelihood method. Using a dynamic probit model, β_y is estimated to be about one, meaning that, as $y_{i,t-1}$ changes from 0 to 1, y^*_{it} changes by one standard deviation (SD): a change of two SD would make $y_{it} = 1$ almost certain. Hyslop also finds $V(\eta_i)/\{V(\eta_i) + V(u_{it})\}$ to be about 0.75 to 0.80. Thus, both state dependence and unobserved heterogeneity matter much. As for ξ,

ξ is estimated to be a small but significant negative number (about -0.22) as in Mühleisen and Zimmermann (1994).

Arulampalam et al. (2000) use the British Household Panel Survey over 1991-1995 for males; a pseudo-related effect is assumed that δ_i is a linear function of $(1/T)\sum_t x_{it}$ plus η_i. Arulampalam et al. deal with the initial value problem as in (iii) above, and they use a two-stage selection correction approach instead of MLE. They find β_y to be 1.05 for men below age 25 and 1.41 for men of age 25 or higher; note that 1.41 SD change in y^*_{it} due to $y_{i,t-1}$ changing from 0 to 1 is quite big. As for $V(\eta_i)/\{V(\eta_i)+V(u_{it})\}$, the estimate varies too much depending on the chosen model, not leading to any firm conclusions.

Phimister et al. (2002) use the Canadian Survey of Labor Income and Dynamics over 1993-1996 for females; unrelated effect is assumed. Phimister et al. deal with the initial value problem as in (iii) above, and they use a two-stage selection correction approach instead of MLE as in Arulampalam et al. (2000). Phimister et al. find $\beta_y = 1.50$ and $V(\delta_i)/\{V(\delta_i)+V(u_{it})\} = 0.39$ (both are significant). That is, state dependence is strong and the unobserved heterogeneity accounts for 39% of the total error term variance.

Knights et al. (2002) use the Australian Longitudinal Survey over 1985-1988 for both males and females; unrelated effect is assumed. They use an unrelated-effect MLE dealing with the initial value problem as in (iii) above; the genuine MLE is used, not the two-stage selection correction approach. β_y ranges from 0.512 for low-educated males, 0.772 for high-educated males, 0.838 for high-educated females, and 0.903 for low-educated females (all significant).

Overall, the following remarks are in order to motivate our study. First, β_y ranges from 0.5 to 1.5 depending on the group of subjects in the data (the estimates in Mühleisen and Zimmermann (1994) are not directly comparable to these numbers, because interaction terms involving $y_{i,t-1}$ are used); this fact along with significant interaction terms involving $y_{i,t-1}$ in Mühleisen and Zimmermann (1994) calls for inclusion of interaction terms between $y_{i,t-1}$ and x_{it}, which is not done in the literature. Second, the (pseudo-) unrelated-effect assumption and the initial value problem in the literature can be avoided using a recent related-effect 'dynamic conditional logit (DCL)' estimator in Honoré and Kyriazidou (2000); the estimator, however, requires $u_{i1},...,u_{iT}$ to be iid, which is thus adopted in the remainder of this paper. Third, most empirical studies are for developed countries; hence, it will be interesting how the empirical findings for developed countries hold up for countries like Korea—an industrial country growing fast while the traditional barriers for women are still strong.

The rest of this paper is organized as follows. Section 2 describes DCL and show that interaction terms between $y_{i,t-1}$ and x_{it} are allowed in the DCL framework; also shown there are the log-likelihood function for dynamic probits and a three-stage algorithm to speed up the convergence of dynamic probits. Section 3 describes the data set. Section 4 presents the empirical findings. Finally, Section 5 concludes.

2. Methodology

In this section, first, we show DCL in Honoré and Kyriazidou (2000). Second, log-likelihood functions for (pseudo-) unrelated-effect dynamic probits are presented. Third, a three-stage algorithm to speed up the dynamic probits is shown. To be coherent with the conditional logit, we may use logistic distribution for u_{it} and δ_i instead of normal distributions. But this would mean deviating from the literature too far. Instead, we will turn DCL estimates to dynamic-probit-comparable estimates by rescaling them properly. Define

$$x_i \equiv (x_{i1}, x_{i2}, ..., x_{iT})'.$$

2.1 Dynamic Conditional Logit (DCL)

Suppose

$$u_{it}\text{'s follow logistic distribution, iid over } t \text{ and } i \text{, and are independent of } y_{i1}, \delta_i, x_i \quad (2.1)$$

To understand the generality as well as the limitations of DCL in Honoré and Kyriazidou (2000), it is necessary to understand the main idea of the estimator. Let

$$\begin{aligned} P(y_{i1} &= 1|x_i, \delta_i) \equiv p_1; \\ P(y_{it} &= 1|x_i, \delta_i, y_{i1}, ..., y_{i,t-1}) \qquad (2.2) \\ &= \exp(\beta_y y_{i,t-1} + x_{it}'\beta + \delta_i)/\{1 + \exp(\beta_y y_{i,t-1} + x_{it}'\beta + \delta_i)\}, \quad for\ t = 2, ..., T; \end{aligned}$$

note that period 1 model is not specified. Consider two events

$$A \equiv \{y_{i1} = d_1, y_{i2} = 0, y_{i3} = 1, y_{i4} = d_4\}\ and\ B \equiv \{y_{i1} = d_1, y_{i2} = 1, y_{i3} = 0, y_{i4} = d_4\};$$

the two events differ only in the middle two variables y_{i2} and y_{i3}.

Observe

$$\begin{aligned}
&P(A|x_i,\delta_i) \\
= \;& p_1^{d_1}(1-p_1)^{1-d_1}\cdot 1/\{1+\exp(\beta_y d_1+x_{i2}'\beta+\delta_i)\}\cdot\exp(x_{i3}'\beta+\delta_i)/\{1+\exp(x_{i3}'\beta+\delta_i)\} \\
& \cdot\exp(d_4(\beta_y+x_{i4}'\beta+\delta_i))/\{1+\exp(\beta_y+x_{i4}'\beta+\delta_i)\},
\end{aligned}$$

which has four terms on the right-hand side. The first term is for y_{i1}, the second is for $(y_{i2}=0)|y_{i1}$, the third is for $(y_{i3}=1)|(y_{i2}=0)$, and the fourth is for $y_{i4}|(y_{i3}=1)$. Analogously, observe

$$\begin{aligned}
P(B|x_i,\delta_i) \;=\;& p_1^{d_1}(1-p_1)^{1-d_1}\cdot\exp(\beta_y d_1+x_{i2}'\beta+\delta_i)/\{1+\exp(\beta_y d_1+x_{i2}'\beta+\delta_i)\} \\
& \cdot 1/\{1+\exp(\beta_y+x_{i3}'\beta+\delta_i)\}\cdot\exp(d_4(x_{i4}'\beta+\delta_i))/\{1+\exp(x_{i4}'\beta+\delta_i)\}.
\end{aligned}$$

Define A_0, A_1, B_0, and B_1 such that $P(A|x_{i2},...,x_{iT},\delta_i)=A_0/A_1$ and $P(B|x_{i2},...,x_{iT},\delta_i)=B_0/B_1$. That is,

$$\begin{aligned}
A_1 &= \{1+\exp(\beta_y d_1+x_{i2}'\beta+\delta_i)\}\ \{1+\exp(x_{i3}'\beta+\delta_i)\}\ \{1+\exp(\beta_y+x_{i4}'\beta+\delta_i)\}, \\
B_1 &= \{1+\exp(\beta_y d_1+x_{i2}'\beta+\delta_i)\}\ \{1+\exp(\beta_y+x_{i3}'\beta+\delta_i)\}\ \{1+\exp(x_{i4}'\beta+\delta_i)\};
\end{aligned}$$

A_1 and B_1 share the same first term. Although $A_1\neq B_1$, if $x_{i3}=x_{i4}$, then the last two terms in A_1 and B_1 agree to result in $A_1=B_1$, which is the key idea. Hence, under $x_{i3}=x_{i4}$, the denominators A_1 and B_1 drop out in the following conditional probability that is free of δ_i:

$$P(A\,|x_i,\delta_i,A\cup B,x_{i3}=x_{i4})=1/\{1+\exp(\beta_y(d_1-d_4)+(x_{i2}-x_{i3})'\beta)\}.$$

The conditioning event $A\cup B$ requires $y_{i2}\neq y_{i3}\Longleftrightarrow y_{i2}+y_{i3}=1$.

When x_{it} is discrete, the DCL log-likelihood function is

$$\begin{aligned}
\sum_i 1[y_{i2}+y_{i3} \;=\;& 1]\ 1[x_{i3}=x_{i4}]\ \ln[\exp\{\beta_y(y_{i1}-y_{i4})+(x_{i2}-x_{i3})'\beta\}^{y_{i2}} \\
& /\{1+\exp(\beta_y(y_{i1}-y_{i4})+(x_{i2}-x_{i3})'\beta)\}]. \qquad (2.3)
\end{aligned}$$

The estimator is $\sqrt{N}$-consistent and asymptotically normal as in other MLE's.

Since δ_i is removed, DCL allows an arbitrary relation between δ_i and x_i. Also, interaction between $y_{i,t-1}$ and x_{it} is allowed. To see this, add $y_{i,t-1}z_{it}'\beta_{yz}$ (β_{yz} is a parameter vector and z_{it} is a subvector of x_{it}) to the last two terms of in A_1 and B_1 to get, respectively,

$$\begin{aligned}
&\{1+\exp(x_{i3}'\beta+\delta_i)\}\cdot\{1+\exp(\beta_y+z_{i4}'\beta_{yz}+x_{i4}'\beta+\delta_i)\}, \\
&\{1+\exp(\beta_y+z_{i3}'\beta_{yz}+x_{i3}'\beta+\delta_i)\}\cdot\{1+\exp(x_{i4}'\beta+\delta_i)\};
\end{aligned}$$

it still holds that $A_1 = B_1$ if $x_{i3} = x_{i4}$. With interaction terms allowed, the log-likelihood function becomes

$$\sum_i 1[y_{i2} + y_{i3} = 1]\ 1[x_{i3} = x_{i4}]\ \ln[\exp\{\beta_y(y_{i1} - y_{i4}) + (y_{i1}z_{i1} - y_{i4}z_{i4})'\beta_{yz} + (x_{i2} - x_{i3})'\beta\}^{y_{i2}}$$
$$/\{1 + \exp(\beta_y(y_{i1} - y_{i4}) + (y_{i1}z_{i1} - y_{i4}z_{i4})'\beta_{yz} + (x_{i2} - x_{i3})'\beta)\}]. \quad (2.4)$$

Although all time-invariant regressors are removed in $x_{i2} - x_{i3}$, they may appear interacting with $y_{i,t-1}$. Since the SD of the logistic distribution is 1.8, we will divide the DCL estimates by 1.8 so that they become comparable to those from dynamic probit; this division does not affect the t-values.

The main restriction of DCL other than (2.1), is $x_{i3} = x_{i3}$, which has a number of consequences. First, if x_{it} is continuous, then a nonparametric smoothing should replace $1[x_{i3} = x_{i4}]$; this slows down the convergence rate of the estimator and makes the estimator bandwidth-dependent. Second, suppose there is a macro shock that changes the intercept. Then DCL is not applicable. To see this, let τ_t denote the intercept in the regression function (pulled out of $x'_{it}\beta$), and rewrite the last two terms of A_1 and B_1 as, respectively,

$$\{1 + \exp(\tau_3 + x'_{i3}\beta + \delta_i)\} \cdot \{1 + \exp(\tau_4 + \beta_y + x'_{i4}\beta + \delta_i)\},$$
$$\{1 + \exp(\tau_3 + \beta_y + x'_{i3}\beta + \delta_i)\} \cdot \{1 + \exp(\tau_4 + x'_{i4}\beta + \delta_i)\}$$

which are no longer the same even if $x_{i3} = x_{i4}$. Third, $x_{i3} = x_{i4}$ disallows regressors such as age or job experience.

One may try to get around the condition $x_{i3} = x_{i4}$ by grouping variables. For continuous variables such as yearly income, essentially this amounts to a nonparametric smoothing, although this nonparametric aspect is often ignored. For variables such as age (and job experience), this will not work, because (2.3) shows that $x_{i2} \neq x_{i3} = x_{i4}$ is needed, which cannot hold for age; by grouping, we would be forcing this on age. If age dummies are used such as $age20$ for being in the 20's or not, $age20_{i3} = age20_{i4}$ will hold mostly other than when the age crosses the boundary point 20 or 30. This crossing will cause a bias, which is likely to be small. If $y_{i,t-1}age20_{it}$ is used, then the condition $x_{i2} \neq x_{i3}$ is fulfilled mostly by the temporal variation of $y_{i,t-1}$. For our empirical analysis with DCL, age dummies will be used.

Despite the restrictions for DCL, however, if we pick two years during which the macroeconomic conditions are relatively stable and if age or job experience effect over one year is small, then these restrictions may not be too binding for DCL.

2.2 Dynamic Probit

For unrelated effect dynamic probit, the main assumptions are that

δ_i is independent of $u_{i1}, ..., u_{iT}$, and $(\delta, u_{i1}, ..., u_{iT})$ is independent of x_i;

$u_{i2}, ..., u_{iT}$ are iid $N(0, \sigma_u^2)$ and independent of u_{i1} that follows $N(0, \sigma_1^2)$; δ_i follows $N(0, \sigma_\delta^2)$.

Define

$$\sigma_v \equiv SD(v_{it}) = SD(\delta_i + u_{it}) \ \forall t = 2, ..., T,$$

and let Φ and ϕ denote the $N(0,1)$ distribution and density function, respectively. Since,

$$\Phi(a)^y \cdot \{1 - \Phi(a)\}^{1-y} = \Phi\{a \cdot (2y - 1)\},$$

dividing the period t latent equation by σ_u and the period 1 equation by σ_1, the log-likelihood function to be used is

$$\sum_i \ln[\int \Phi\{(x_1'\alpha/\sigma_1 \ + \ (\delta/\sigma_\delta) \cdot \beta_\delta \sigma_\delta/\sigma_1)(2y_{i1} - 1)\} \tag{2.5}$$

$$\prod_{t=2}^{T} \Phi\{(y_{i,t-1}\beta_y/\sigma_u \ + y_{i,t-1}z_{it}'\beta_{yz}/\sigma_u \ + x_{it}'\beta/\sigma_u \ + (\delta/\sigma_\delta)\sigma_\delta/\sigma_u)(2y_{it} - 1)\}\phi(\delta/\sigma_\delta)\sigma_\delta^{-1} d\delta \ ]$$

where only x_1 is used in the y_{i1} equation; if desired, x_i can be used without difficulty. Since $\zeta \equiv \delta/\sigma_\delta$ is $N(0,1)$, rewrite the log-likelihood as

$$\sum_i \ln[\int \Phi\{(x_1'\alpha/\sigma_1 \ + \ \zeta\beta_\delta\sigma_\delta/\sigma_1)(2y_{i1} - 1)\} \tag{2.6}$$

$$\prod_{t=2}^{T} \Phi\{(y_{i,t-1}\beta_y/\sigma_u \ + y_{i,t-1}z_{it}'\beta_{yz}/\sigma_u \ + x_{it}'\beta/\sigma_u \ + \zeta\sigma_\delta/\sigma_u)(2y_{it} - 1)\}\phi(\zeta)d\zeta.$$

The identified parameters are

$$\alpha/\sigma_1, \ \beta_\delta\sigma_\delta/\sigma_1, \ \ \beta_y/\sigma_u, \ \beta_{yz}/\sigma_u, \ \beta/\sigma_u, \ \sigma_\delta/\sigma_u; \tag{2.7}$$

the last term σ_δ/σ_u shows how important δ_i is relative to u_{it}.

For pseudo-related effect dynamic probit, consider

$$\delta_i = x_{i1}'\mu_1 +, ..., +x_{iT}'\mu_T + \eta_i, \quad \text{or a simpler version} \quad \delta_i = \bar{x}_i'\mu + \eta_i.$$

Since the former is computationally too demanding, although it can be dealt with in principle as done in Lee (2002), we adopt the simpler version that includes the restriction $\mu_1 =, ..., = \mu_T$. Substitute $\delta_i = \bar{x}_i'\mu + \eta_i$ into (2.5) to get

$$\sum_i \ln[\int \Phi\{(x_1'\alpha/\sigma_1 \ + \ (\bar{x}_i'\mu + \eta_i)\beta_\delta/\sigma_1)(2y_{i1} - 1)\} \tag{2.8}$$

$$\prod_{t=2}^{T} \Phi\{(y_{i,t-1}\beta_y/\sigma_u \ + y_{i,t-1}z_{it}'\beta_{yz}/\sigma_u \ + x_{it}'\beta/\sigma_u \ + (\bar{x}_i'\mu + \eta_i)/\sigma_u)(2y_{it} - 1)\}\phi(\eta/\sigma_\eta)\sigma_\eta^{-1}d\eta \ ].$$

With $\zeta = \eta/\sigma_\eta$, this can be rewritten as

$$\sum_i \ln[\int \Phi\{(x_1'\alpha/\sigma_1 \ + \ \bar{x}_i'\mu\beta_\delta/\sigma_1 \ + \ \zeta\beta_\delta\sigma_\eta/\sigma_1)(2y_{i1} - 1)\} \tag{2.9}$$

$$\prod_{t=2}^{T} \Phi\{(y_{i,t-1}\beta_y/\sigma_u \ + y_{i,t-1}z_{it}'\beta_{yz}/\sigma_u \ + x_{it}'\beta/\sigma_u \ + \bar{x}_i'\mu/\sigma_u \ + \zeta\sigma_\eta/\sigma_u)(2y_{it} - 1)\}\phi(\zeta)d\zeta.$$

The identified parameters are (compare to (2.7))

$$\alpha/\sigma_1, \ \mu\beta_\delta/\sigma_1, \ \beta_\delta\sigma_\eta/\sigma_1, \ \ \beta_y/\sigma_u, \ \beta_{yz}/\sigma_u, \ \beta/\sigma_u, \ \mu/\sigma_u, \ \sigma_\eta/\sigma_u. \tag{2.10}$$

In practice, although x_{it} includes both time-constant and time-variant variables, $\bar{x}_i$ should consist only of time-variants; otherwise the time-constant variables would be used twice as regressors in the same equation. This means that the coefficients for the time-constant variables in x_{it} are in fact the sum of the actual coefficients plus those for $\bar{x}_i$.

If we treat y_{i1} as fixed, then we just have to drop the first period likelihood component in (2.6) or in (2.9). The fixed y_{i1} assumption is often used in practice; also, comparing the estimates under fixed y_{i1} to those under random y_{i1} will show how sensitive the dynamic probit is to the initial period.

In short, we will examine five sets of estimates in our empirical section later:

1. DCL: (2.4).

2. unrelated-effect dynamic probit with y_{i1} fixed: (2.6) with period 1 removed.

3. unrelated-effect dynamic probit with y_{i1} random: (2.6).

4. pseudo-related-effect dynamic probit with y_{i1} fixed: (2.9) with period 1 removed.

5. pseudo-related-effect dynamic probit with y_{i1} random: (2.9).

Integration for dynamic probit can be done numerically. But this turned out to be too time-consuming. Instead, we use Monte Carlo integration with the number of random draws being 10 to save time. As to be shown later, although 10 may sound small, the final results change little when the number of random draws increases to 35. In principle, the asymptotic theory for simulated log-likelihood estimation calls for the random draw number to be infinite. With the number 10, the dynamic probit programs took about two hours on a pentium four PC; with 35, it took more than a day. Hence, with 35, we use the following three-stage algorithm to speed up the program.

2.3 Three-Stage Algorithm to Speed Up Dynamic Probit

The integration in (2.9) can be done either numerically or with a Monte Carlo simulation. Either way, maximizing (2.9) can be rather time-consuming. But there is a three-stage algorithm that pretty much halves the computation time. If the computation takes a minute, halving it would not be a big deal, but if the computation takes several hours as in our case, then halving it would be helpful.

The idea goes as follows. First, estimate the initial period parameters with a simple probit. Second, estimate the remaining parameters while keeping the first-stage parameters intact. Third, take one "Newton-Raphson" step from the second-stage estimates. The estimator obtained this way is asymptotically as efficient as the single-stage MLE. In practice, however, the third stage may be iterated until convergence instead of taking just one step. A complicating factor for this three stage idea is that δ is not conditioned on in the initial period equation at the first stage; this is explained in the following.

Since the first period error term without the normalization by σ_1 is $\beta_\delta \eta + u_1$ in the pseudo-related effect, define

$$\sigma_a^2 \equiv V(\beta_\delta \eta + u_1) = \beta_\delta^2 \sigma_\eta^2 + \sigma_1^2 \implies (\beta_\delta \sigma_\eta / \sigma_a)^2 = 1 - (\sigma_1/\sigma_a)^2. \tag{2.11}$$

In the first-stage probit, since the error term is $\beta_\delta \eta + u_1$ when δ is not conditioned on, σ_a is the right normalizing scale factor, differently from σ_1 appearing in (2.9) where δ is first conditioned upon before finally integrated out. The first stage probit yields $\widehat{\alpha/\sigma_a}$ and $\mu\widehat{\beta_\delta/\sigma_a}$.

In the second stage, we need to maximize (2.9) with the first stage estimates plugged in. For this, rewrite the initial period likelihood contribution in (2.9) as

$$\Phi\{(x_1'(\alpha/\sigma_a)(\sigma_a/\sigma_1) \;+\; \bar{x}_i'(\mu\beta_\delta/\sigma_a)(\sigma_a/\sigma_1) \;+\; \zeta(\beta_\delta\sigma_\eta/\sigma_a)(\sigma_a/\sigma_1))\cdot(2y_{i1}-1)\}$$

$$\simeq \;\; \Phi\{\;[x_1'\widehat{\alpha/\sigma_a}(\sigma_a/\sigma_1) \;+\; \bar{x}_i'\widehat{\mu\beta_\delta/\sigma_a}(\sigma_a/\sigma_1) \;+\; \zeta\{1-(\sigma_a/\sigma_1)^{-2}\}^{1/2}(\sigma_a/\sigma_1)]\cdot(2y_{i1}-1)\;\};$$

only σ_a/σ_1 (>1) needs to be estimated. Thus, the estimated parameters in the second-stage are

$$\sigma_a/\sigma_1,\; \beta_y/\sigma_u,\; \beta_{yz}/\sigma_u,\; \beta/\sigma_u,\; \mu/\sigma_u,\; \sigma_\eta/\sigma_u. \tag{2.12}$$

In the third stage, combine the first-stage probit and the second-stage estimates to get (2.10), to take one Newton-Raphson step from the estimates; better yet, the third stage can be iterated fully as mentioned above.

3. Data and Descriptive Statistics

Our data set consists of the first four waves (1998-2001) of the Korea Labor and Income Panel Study (KLIPS). Excluding the dropouts, the data set is balanced with $N =$ 3882 for women of age 15-65.

Other than the response variable y_{it} for work or not in year t, the following regressors are used: *age*, the number of children aged 1-3 (*ch1*), 4-7 (*ch2*), and 8 -13 (*ch3*), six education dummies for the final education completion level being primary school (*ed1*), middle school (*ed2*), high school (*ed3*), junior college (*ed4*), college (*ed5*), and MA degree or higher (*ed6*), *inc* for the logarithm of household income other than the female's own income, *mar* for married or not, *jtr* for any job training or not, and *cert* for the number of certificates (for job-related skills). From age, *age2* (= $age^2/100$), *age20* for being in 20's or not, *age30* for being in 30's or not, and *age40* for being in 40's or above are also used as regressors. Ed1 is the base education case and thus not used in estimation.

Table 1: Descriptive Statistics

variable	mean	SD	min	med	max
y	0.457	0.498	0	0	1
age	37.510	12.871	15	37	64
age20	0.210	0.408	0	0	1
age30	0.270	0.444	0	0	1
ch1	0.086	0.294	0	0	3
ch2	0.099	0.309	0	0	2
ch3	0.197	0.475	0	0	3
ed1	0.142	0.350	0	0	1
ed2	0.189	0.392	0	0	1
ed3	0.466	0.499	0	0	1
ed4	0.063	0.268	0	0	1
ed5	0.078	0.268	0	0	1
ed6	0.006	0.075	0	0	1
cert	0.256	0.729	0	0	8
income	3,673	166,772	0	1109	12,000,000
jtr	0.036	0.187	0	0	1
mar	0.752	0.432	0	1	1

Table 1 shows the descriptive statistics over the four years. The labor participation rate is still relatively low (46%) in Korea. The high-school graduation is the majority. Income in Table 1 is not in log, but in Korean-Won$\times$10000, which is roughly \$8; thus, the median household income other than the female's own income is $\$8872 = 1109 \times 8$. About 75% are married.

Table 2 shows age-group breakdown of workers for each year; in the last row, the numbers in parentheses are the number of working women in our data. During 1999-2001, the shares of teens and 30's have declined; the shares of 40's and 60-65 have increased steadily; the shares of 20's and 50's stayed about the same. Forming two groups, 30's or below and 40's or above, the first group's share has declined (55 to 52 to 49) during 1999-2001, whereas the second group's share has risen.

Table 2: Age Group Percentage among Workers for Each Year

age bracket	1998	1999	2000	2001
15~19	2	3	2	1
20~29	20	22	22	23
30~39	32	30	28	25
40~49	28	27	29	31
50~59	15	14	14	14
60~65	3	4	5	6
sum	100 (1560)	100 (1877)	100 (1813)	100 (1902)

Ignore all regressors for a while and consider $y_{it} = 1[\beta_y y_{i,t-1} + v_{it} \geq 0]$, $v_{it} \sim N(0,1)$. If $\beta_y = 0$, then y_{it} will take 0 or 1 randomly, depending on whether v_{it} is negative or positive. Suppose $\beta_y = 0.5$, which plays no role if $y_{i,t-1} = 0$. But once $y_{i,t-1} = 1$, then $y_{it} = 1[0.5 + v_{it} \geq 0]$: it will take $v_{it} < -0.5$ for the woman not to work at t. If $\beta_y = 2$ and $y_{i,t-1} = 1$, then $y_{it} = 1[2 + v_{it} \geq 0]$: it will take $v_{it} < -2$ for the woman not to work at t, the probability of which is very low. In short, once an woman starts to work (maybe because $v_{i,t-1}$ takes a positive value), the stronger the state dependence is, the less likely she goes back to non-working. With $x'_{it}\beta$ in, we get $y_{it} = 1[\beta_y y_{i,t-1} + x'_{it}\beta + v_{it} \geq 0]$, which puts female i in a less ($x'_{it}\beta < 0$) or more ($x'_{it}\beta > 0$) favorable position to work; other than this aspect, an analogous interpretation can be given to β_y.

Table 3 shows the percentage of labor participation status changes over the four year period. For instance, during 1998-1999, the proportion of 0→1 in 0→0 or 0→1 is 14/60=0.23 whereas the proportion of 1→1 in 1→0 or 1→1 is 34/40=0.85. This suggests a strong state dependence, but without x_{it} controlled for, nothing definite can be said, because the group 0→0 or 0→1 may have smaller $x'_{it}\beta$ than the other group.

Table 3: Change in Labor Participation

years	0→0	0→1	1→0	1→1	sum
1998-1999	46	14	6	34	100
1999-2000	45	7	9	39	100
2000-2001	45	9	6	40	100

It is well known that income variables are not trustworthy in household surveys, but in our data, we found out education is also error-ridden: many women's education level have decreased over the years! We corrected these cases by using the minimum eduction level reported: for examle, if a woman reports college education in wave 1 but only highschool education in wave 2, then we set the education for wave 1 at highschool education. The following shows the percentage of corrections we had to make over the total $N \times T$ ($15528 = 3882 \times 4$) observations:

	ed1	ed2	ed3	ed4	ed5	ed6
percentage corrected	23.4	28.0	48.5	11.6	14.2	1.1
number of observations	3636	4342	7537	1802	2197	169

For instance, the number of differences before correction and after correction in ed1 is 3636, which is 23.4% of 15528. In addition to genuine errors in data collection, we suspect deliberate misrepresentation by the respondents, with the Korean people being very conscious of their education levels. Our way of correction using the minimum reported education level may not be a good one, but without an external validation sample, there is no way to see which correction method is the best. Although it is unfornatute that this much correction had to be made, we proceed with the corrected data.

4. Empirical Findings for Korean Female Labor Participation

Table 4 shows the estimation results for DCL. To make the estimates comparable to the dynamic probit estimates to appear later, we divided the DCL estimates by 1.8 and put them in the last column of Table 4. We will base our interpretation on those "normalized" estimates; division by 1.8 does not affect the t-values.

Before we proceed further, two remarks are in order. First, although we tried to include as many variables to be used for dynamic probit as possible in Table 4, the temporal variation in some variables is too small for the coefficients to be estimable; hence, those variables are excluded from Table 4. Second, income is not used for DCL, because it is a continuous variable requiring smoothing in DCL if used; although income turns out to be statistically significant for dynamic probits, its coefficient is about -0.2 or smaller in magnitude. Thus, the bias due to omitting income is likely to be small.

For Table 4, we tried many interaction terms with y_{t-1} and only ed4 came out some significance; mar, age20, age30, and age40 are included for later comparisons with dynamic probits. Other than for ed4=1, the state dependence is 0.679; 0.679 (with the t-value 1.51) falls close to the low end of the state dependence numbers seen in the literature for developed countries. For women with junior college degree (ed4=1), the state dependence is 2.28, which means almost no change in work status. This may be owing to the fact that the education at junior college tends to be job-oriented and job-specific. For women in 20's, the state dependence is almost zero: $0.106 = 0.679 - 0.573$. The state dependence for women in 30's is also small: $0.161 = 0.679 - 0.518$. Other than state dependence, ch2, and jtr have fairly significant coefficients, comparable to that of y_{t-1} in magnitude: ch2 decreases labor participation whereas job-training increases it as one would expect. About 5% of the women in our data change ed3 (from 0 to 1) and about 1% change in ed5, and probably as the consequece, ed3 and ed5 have insignificant estimates.

Table 4: Dynamic Conditional Logit

	estimate	(t-value)	estimate/1.8
y_{t-1}	1.222	(1.51)	0.679
y_{t-1}*mar	0.284	(0.50)	0.158
y_{t-1}*ed4	2.882	(2.66)	1.601
y_{t-1}*age20	-1.032	(-1.17)	-0.573
y_{t-1}*age30	-0.932	(-0.95)	-0.518
y_{t-1}*age40	-0.058	(-0.06)	-0.032
ch1	0.751	(1.15)	0.417
ch2	-1.008	(-1.59)	-0.560
ch3	-0.240	(-0.57)	-0.133
ed3	0.106	(0.12)	0.614
ed5	0.201	(0.17)	0.112
age20*ed3	-0.901.	(-1.47)	-0.501
age30*ed3	-1.112	(-1.27)	-0.617
cert	-0.095	(-0.08)	-0.053
jtr	1.227	(1.79)	0.682
conditional log-like.		-238.111	

Table 5 presents unrelated-effect dynamic probit results. The columns with "y_1-fixed" are vastly different from the columns with "y_1-random". The reason is clear: the middle columns for the initial period have many significant estimates in terms of magnitude and t-value. Omitting the initial period equation introduces huge biases for the estimator treating y_1 as fixed; this estimator should not be used. We will base our interpretation of unrelated-effect dynamic probit on the last two columns of Table 5; the estimates for the initial period are of 'reduced-form type', and as such, they are difficult to interpret.

The magnitude of y_{t-1} for dynamic unrelated probit is slightly smaller than that for DCL, but the estimate for y_{t-1}∗ed4 is about five times smaller and insignificant. The estimate for y_{t-1}∗mar is significant with the magnitude greater than that for y_{t-1}. The estimates for y_{t-1}∗age20 and y_{t-1}∗age30 are not too far from those for DCL: state dependence is almost zero for single women in 20's and 30's. Age has the usual up and down pattern for labor participation. The number of young children matters, but judging from the relative magnitude, not as much as y_{t-1} and its interaction terms do.

Ed3, ed4 and ed5 are negative, but this is due to the interaction terms between age dummies and education levels. For instance, the effect of ed5 on labor participation is

$$\begin{aligned} &\text{for women in 20's:} && 1.390 = 2.227 - 0.837 \\ &\text{for women in 30's:} && 0.318 = 1.155 - 0.837 \\ &\text{for women in 40's or above:} && -0.837 \end{aligned}$$

That is, for mature women, college education is a hindrance for labor participation.

The estimate for cert is significant but small. If income increases by 100%, then this will decrease labor participation propensity by 0.031∗SD(error), which is quite small. Job-training has a significant estimate (0.516) whose magnitude is close to that for DCL (0.682). Marriage with the estimate -1.327 seems to be most detrimental to female labor participation. The ratio of the time-invariant error SD to time-variant error SD is about 1.344; that is, the ratio of time-invariant error variance to the total error variance is about $0.643 = 1.344^2/(1.344^2 + 1)$, which is close to the estimates seen in the literature.

Table 5: Unrelated-Effect Dynamic Probit						
	y_1 fixed: est. (tv)		y_1 random: est. (tv)			
y_{t-1}	0.099	(2.88)			0.511	(2.29)
y_{t-1}*mar	0.042	(2.01)			0.688	(4.36)
y_{t-1}*ed4	-0.531	(-6.09)	* for initial period *		0.300	(1.55)
y_{t-1}*age20	0.416	(4.66)			-0.481	(-2.07)
y_{t-1}*age30	-0.026	(-5.30)			-0.474	(-1.82)
y_{t-1}*age40	-0.093	(-0.45)			-0.003	(-0.01)
one	1.584	(9.98)	-7.603	(-13.58)	-6.147	(-13.57)
age	0.457	(4.68)	0.428	(13.80)	0.349	(13.25)
age2	0.049	(0.41)	-0.476	(-13.28)	-0.399	(-13.45)
ch1	0.243	(3.74)	-0.329	(-2.98)	-0.238	(-2.84)
ch2	0.945	(4.15)	-0.039	(-0.39)	-0.052	(-0.73)
ch3	1.119	(6.86)	0.100	(1.39)	0.096	(1.79)
ed3	-0.323	(-1.89)	-0.223	(-2.07)	-0.254	(-3.10)
ed4	-0.253	(-1.37)	0.201	(0.48)	-0.666	(-1.77)
ed5	0.118	(0.63)	-0.601	(-1.90)	-0.837	(-3.85)
ed6	-2.418	(-10.93)	1.048	(1.98)	0.626	(1.28)
age20*ed3	0.106	(8.39)	0.544	(3.90)	0.252	(2.51)
age20*ed4	-0.126	(-8.81)	0.990	(2.13)	1.837	(4.74)
age20*ed5	-0.107	(-2.19)	1.847	(5.01)	2.227	(8.32)
age30*ed3	-0.325	(-1.45)	0.019	(-0.15)	0.135	(1.43)
age30*ed4	-0.369	(-2.77)	-0.711	(-1.50)	0.223	(0.58)
age30*ed5	0.113	(0.42)	0.683	(1.90)	1.155	(4.78)
cert	0.128	(2.12)	0.282	(5.39)	0.163	(4.07)
inc	-0.150	(-2.54)	-0.082	(-6.32)	-0.031	(-4.94)
jtr	0.533	(3.39)	0.027	(0.24)	0.516	(4.30)
mar	0.249	(1.05)	-1.104	(-6.71)	-1.327	(-8.44)
σ_δ/σ_u	0.005	(0.09)	1.291	(16.99)	1.344	(15.15)
log-likelihood	-4966.47		-7080.355			

Table 6: Pseudo-Related-Effect Dynamic Probit									
	y_1 fixed: est. (tv)				y_1 random: est. (tv)				
y_{t-1}	0.195	(2.44)			0.571	(2.57)			
y_{t-1}*mar	-0.206	(-1.51)			0.664	(4.18)			
y_{t-1}*ed4	0.305	(1.51)	* for initial period *		0.327	(1.67)			
y_{t-1}*age20	0.883	(5.45)			-0.538	(-2.31)			
y_{t-1}*age30	-0.062	(-4.62)			-0.547	(-2.09)			
y_{t-1}*age40	-0.062	(-0.30)			-0.087	(-0.33)			
one	1.524	(9.40)	-5.828	(-10.19)	-4.212	(-9.58)			
age	0.506	(5.10)	0.356	(11.10)	0.289	(11.13)			
age2	0.086	(0.72)	-0.400	(-10.85)	-0.335	(-11.50)			
ch1	0.320	(4.88)	-0.216	(-1.55)	0.030	(0.30)	$\overline{ch1}$	-1.289	(-6.54)
ch2	1.110	(4.69)	0.045	(0.35)	-0.002	(-0.03)	$\overline{ch2}$	-0.286	(-1.68)
ch3	1.279	(7.68)	0.165	(1.52)	-0.135	(-1.67)	$\overline{ch3}$	0.318	(2.76)
ed3	-0.336	(-1.93)	-0.169	(-1.54)	-0.201	(-2.42)			
ed4	-0.305	(-1.62)	0.235	(0.54)	-0.624	(-1.60)			
ed5	0.104	(0.55)	-0.594	(-1.84)	-0.809	(-3.62)			
ed6	-2.143	(-9.40)	0.793	(1.14)	1.541	(3.25)			
age20*ed3	0.105	(8.27)	0.558	(4.00)	0.356	(3.49)			
age20*ed4	-0.127	(-8.90)	1.045	(2.20)	1.988	(4.88)			
age20*ed5	-0.136	(-2.80)	2.009	(5.32)	2.411	(8.80)			
age30*ed3	-0.399	(-1.71)	0.041	(0.31)	0.187	(1.92)			
age30*ed4	-0.419	(-3.06)	-0.640	(-1.31)	0.342	(0.84)			
age30*ed5	0.189	(0.71)	0.848	(2.34)	1.425	(5.81)			
cert	0.178	(2.90)	0.276	(1.56)	0.240	(1.27)	$\overline{cert}$	-0.089	(-0.45)
inc	0.106	(1.35)	-0.079	(-4.65)	-0.025	(-3.75)	$\overline{inc}$	-0.204	(-7.37)
jtr	0.703	(4.41)	0.051	(0.28)	0.298	(2.40)	$\overline{jtr}$	2.030	(6.16)
mar	0.362	(1.47)	-1.391	(-3.60)	-1.485	(-4.46)	$\overline{mar}$	0.656	(2.07)
σ_η/σ_u	-0.039	(-0.37)	1.342	(17.53)	1.395	(15.42)			
log-likelihood	-4915.768				-6988.534				

Table 6 presents pseudo-related-effect dynamic probit results. As in Table 5, the columns with "y_1-fixed" are vastly different from the columns with "y_1-random": omitting the initial period equation introduces huge biases. As in Table 5, we will base our interpretation on the columns for "y_1-random". The last two columns of Table 6 show the estimates and t-values of the averaged time-variant regressors for pseudo-related effect; to save space, the corresponding estimates for the "y_1-fixed" case and for the initial period in the "y_1-random" case are omitted. The coefficients for averaged time-variant regressors can be usually interpreted as effects of permanent (or long-term) change; for example, the estimate for $\overline{inc}$ is the effect of the 'permanent income' which is time-constant, whereas the estimate for inc is the effect of 'transitory income' which is time-variant.

The estimates for y_{t-1} and its interaction terms are little different from those in Table 5, and the comments made for Table 5 apply to Table 6 with little change; the same is true of ed3, ed4, and ed5, and their interaction terms with age20 and age30. Age has again the usual up and down pattern for labor participation.

The estimate for $ch1$ is almost zero, for the effects are picked up by $\overline{ch1}$: $\overline{ch1}$ has a significant negative coefficient that is also large in magnitude (more than twice the effect of y_{t-1}). Interpreting $\overline{ch1}$ as a permanent or long-term component is somewhat misleading: $\overline{ch1}$ has a "permanent" component only because T is short; if T is 20, then $\overline{ch1}$ should be almost zero.

If "permanent" income $\overline{inc}$ increases by 100%, then this decreases labor participation propensity by 0.204∗SD(error), which is much bigger than in Table 5 but still small in its magnitude; the effect of transitory income is eight times smaller.

The permanent component $\overline{jtr}$ of job-training has a significant estimate (2.03) which is also much larger than the estimates (0.516) for the unrelated-effect probit; the magnitude is more than three time the estimate for y_{t-1} (0.571). The transitory component estimate (0.298) is significant but about seven times smaller in magnitude than the permanent component estimate. As in $\overline{ch1}$, it is a little misleading to interpret $\overline{jtr}$ as a permanent component of job-training, because $\overline{jtr}$ would be zero for a large T. Instead, $\overline{jtr}$ should be viewed as eagerness to work that is a time-constant unobserved trait, whereas jtr is the effect of taking a job-training newly. The pseudo- related-effect model captures this feature correctly as the estimate for jrt (0.298) and $\overline{jtr}$ (2.03) show. In the unrelated-effect model for Table 5, $\overline{jtr}$ is omitted, and this omitted variable causes a bias resulting in the estimate for jtr being 0.516 that falls between 0.298 and 2.03. The DCL estimate 0.682 for jtr is much bigger than

0.298 in Table 6, suggesting that the controlling for the eagerness to work with $\overline{jtr}$ may be inadequate.

Interestingly, the long-term marriage effect from $\overline{mar}$ (i.e., the effect of *being married*) is positive and significant; the magnitude is slightly greater than that for state dependence. In contrast, the transitory component mar has a big negative effect (-1.485), which is the effect of *getting married*. In the short-run, marriage is the most detrimental for female labor participation.

The ratio σ_η/σ_u is similar to the ratio σ_δ/σ_u. This may look strange, because supposedly time-invariant components have been pulled out of δ, which would mean a smaller SD of time-constant error. For this, recall the initial "motivating" equation for related-effect: $\delta_i = x'_{i1}\mu_1 +, ..., +x'_{iT}\mu_T + \eta_i$. This shows that, part of u_{it} related to $x_{i1}, ..., x_{iT}$ may have been taken out as well. That is, σ_u can also decrease in the pseudo-related-effeect probit.

Finally, Table 7 presents the three-stage estimation results. Since this takes much less time, as already mentioned, we could afford to increase the random draw number from 10 to 35; also, instead of taking one Newton-Raphson step in the last stage, we did a full iteration. This "finer" estimation effort resulted in the log-likelihood value increasing slightly from -6988.534 in Table 6 to -6986.907 in Table 7. Overall, Table 6 and Table 7 are not much different, and the comments made for Table 6 apply to Table 7 as well; the only exception seems to be y_{t-1}ed4, which is now significant with a slightly bigger estimate.

Table 7: Pseudo-Related-Effect Dynamic Probit (three stage)						
			y_1 random: est. (tv)			
y_{t-1}			0.538	(3.46)		
y_{t-1}*mar			0.663	(4.84)		
y_{t-1}*ed4	* for initial period *		0.415	(2.14)		
y_{t-1}*age20			-0.580	(-3.90)		
y_{t-1}*age30			-0.527	(-4.84)		
y_{t-1}*age40			-0.055	(-0.52)		
one	-5.994	(-10.42)	-4.339	(-9.65)		
age	0.357	(11.23)	0.292	(11.19)		
age2	-0.405	(-11.07)	-0.346	(-11.49)		
ch1	-0.258	(-1.84)	0.029	(0.29)	$\overline{ch1}$	-1.449 (-6.67)
ch2	0.037	(0.29)	-0.006	(-0.07)	$\overline{ch2}$	-0.327 (-1.88)
ch3	0.153	(1.41)	-0.126	(-1.56)	$\overline{ch3}$	0.252 (2.22)
ed3	-0.227	(-2.12)	-0.248	(-3.07)		
ed4	0.189	(0.48)	-0.681	(-1.76)		
ed5	-0.581	(-1.87)	-0.817	(-3.58)		
ed6	0.787	(1.18)	1.416	(3.49)		
age20*ed3	0.618	(4.37)	0.341	(3.31)		
age20*ed4	1.117	(2.55)	1.982	(4.92)		
age20*ed5	1.990	(5.46)	2.387	(8.70)		
age30*ed3	0.103	(0.77)	0.229	(2.36)		
age30*ed4	-0.609	(-1.35)	0.363	(0.88)		
age30*ed5	0.728	(2.03)	1.317	(5.25)		
cert	0.276	(1.56)	0.237	(1.28)	$\overline{cert}$	-0.095 (-0.49)
inc	-0.079	(-4.69)	-0.025	(-3.84)	$\overline{inc}$	-0.191 (-7.23)
jtr	0.009	(0.05)	0.297	(2.42)	$\overline{jtr}$	1.914 (6.30)
mar	-1.139	(-3.49)	-1.440	(-4.55)	$\overline{mar}$	0.722 (2.33)
σ_η/σ_u	1.291	(17.79)	1.416	(16.02)		
log-likelihood			-6986.907			

5. Conclusions

In this paper, we estimated dynamic labor participation models for Korean women, using a number of models ranging from unrelated-effect, pseudo-related effect, to related-effect models. Despite the differences in the models, we obtained more or less coherent results across the models, and the following findings emerged. Lagged response y_{t-1} matters much with unobserved heterogeneity allowed for. Also, y_{t-1} interacts with a number of variables: marriage dummy, junior (technical) college dummy, and age group dummies for 20's and 30's. The state dependence is about 0.6, and 0.2 to 0.7 higher for married, about 0.5 to 0.6 smaller for women in 20's or 30's; thus, almost no state dependence for single young women. The level of state dependence falls near the low end of the state dependence estimates seen in the literature for developed countries. Junior college education increases state dependence at least by 0.3. Job-training and college education increase labor participation, but college education can be a hindrance for women in their 40's or higher. The transitory effect of getting married is highly negative (-1.4) whereas the long-term effect of being married is positive (about 0.7). These findings lead to policy implications: junior college education more geared for job skills be supported, and policy efforts to increase female labor participation be directed at women of relatively high age. On the methodological front, we suggested a three-stage estimator to reduce the computation time in dynamic probits, and showed that the estimators treating the initial period response as fixed are highly biased and should not be used.

REFERENCES

Arulampalam, W., A.L. Booth, and M.P. Taylor, 2000, Unemployment persistence, Oxford Economic Papers 52, 24-50.

Chamberlain, G., 1984, Panel data, in Handbook of Econometrics II, edited by Z. Griliches and M. Intrilligator, Amsterdam, North Holland.

Heckman, J.J., 1981, The incidental parameters problem and the problem of initial conditions in estimating a discrete time discrete data stochastic process, in Structural analysis of discrete data with econometric applications, edited by C.F. Manski and D. McFadden, MIT Press.

Honoré, B.E. and E. Kyriazidou, 2000, Panel data discrete choice models with lagged dependent variables, Econometrica 68, 839-874.

Hyslop, D.R., 1999, State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women, Econometrica 67, 1255-1294.

Knight, S., M. Harris, and J. Loundes, 2002, Dynamic relationships in the Australian labour market: heterogeneity and state dependence, Economic Record 78, 284-298.

Lee, M.J., 2002, Panel data econometrics: method-of-moments and limited dependent variables, Academic Press.

Mühleisen, M. and K.F. Zimmermann, 1994, A panel analysis of job changes and unemployment, European Economic Review 38, 793-801.

Phimister, E., E. Vera-Toscano, and A. Weersink, 2002, Female participation and labor market attachment in rural Canada, American Journal of Agricultural Economics 84, 210-221.

Shaw, K., 1994, The persistence of female labor supply, Journal of Human Resources 29, 348-378.