2-2005

Analysis of Labor Participation Behavior of Korean Women with Dynamic Probit and Conditional Logit

Myoung-jae LEE
Singapore Management University, mjlee@smu.edu.sg

Yoon-Hee TAE
Korea Development Institute

DOI: https://doi.org/10.1111/j.1468-0084.2005.00110.x

Published in Oxford Bulletin of Economics and Statistics, 2005, 67 (1), 71-91
https://doi.org/10.1111/j.1468-0084.2005.00110.x

Analysis of Labour Participation Behaviour of Korean Women with Dynamic Probit and Conditional Logit*

MYOUNG-JAE LEE† and YOON-HEE TAE‡

†*School of Economics and Social Sciences, Singapore Management University, Singapore (e-mail: mjlee@smu.edu.sg)*
‡*Public Finance and Social Development Division, Korea Development Institute, Chongyangri, Seoul, South Korea (e-mail: happytae@kdi.re.kr)*

Abstract

We analyse the dynamic labour participation behaviour of Korean women. State dependence under unobserved heterogeneity is considered, where the heterogeneity may be unrelated, pseudo-related, or arbitrarily related to regressors. Three minor methodological contributions are made: interaction terms with lagged response are allowed in dynamic conditional logit; a three-stage algorithm for dynamic probit is proposed; and treating the initial response as fixed is shown to be ill-advised. The state dependence is about $0.6 \times$ SD(error), higher for the married or junior college-educated, and lower for women in their twenties and thirties. While education increases participation, college education has negative effects for women in their forties or above. Marriage has a high negative short-term effect but a positive long-term effect.

I. 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 the future. For female labour supply, this means that once a

*This research has been funded by the Wharton-SMU Research Center, Singapore Management University. The authors are grateful to the editor and a referee for their helpful comments.
JEL Classification numbers: C33, C35, J21.

woman makes the choice of working (vs. non-working), the simple fact that she works now will increase the likelihood that she works in the future, with the other things held constant. In this case, if the government wants to increase female labour supply, then the policy effort should be directed to increasing early participation rates rather than 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 onto 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

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

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 state dependence.

State dependence in female labour supply can arise for a number of reasons. First, preference for work (vs. leisure) can be intertemporally non-separable: leisures at two time points can be complements or substitutes. Secondly, 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. Thirdly, 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. Fourthly, work status may have a signalling (or scarring) effect about productivity of the female. Taken together, state dependence can be positive or negative. But all the literature points to positive effects of substantial magnitude. Just like most things in life, temporal behaviour 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:

$$\beta_y \neq 0, \quad \delta_i \neq 0, \quad \text{and} \quad \text{COR}(u_{i,t-1}, u_{it}) \neq 0.$$

Choice persistence can also result from persistence in x_{it}, but this will be ignored because x_{it} can be controlled for. For reasons given later, we will consider only the first two sources for our empirical work – state dependence and unobserved heterogeneity – assuming that $u_{i1}, \dots, u_{iT}$ are independent.

Unobserved heterogeneity has a different implication from state dependence on policies: a policy under unobserved heterogeneity does not work unless it changes the individual heterogeneity; this would require more long-term policy outlay instead of one shot that might be sufficient under state dependence. Separating state dependence from unobserved heterogeneity cannot be done with cross-section data or 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 $x_{i1}, \dots, x_{iT}$ arbitrarily, and the other is 'random effect' where δ_i is independent of $x_{i1}, \dots, x_{iT}$. 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 studies dealing with state dependence for labour 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 modelling δ_i as a linear function of $x_{i1}, \dots, x_{iT}$ 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 $\alpha_1, \dots, \alpha_T$ and α_δ, the three approaches are

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

(ii)

$$y_{i1} = 1[x'_{i1}\alpha_1 + \cdots + x'_{iT}\alpha_T + v_{i1} > 0], \qquad \text{COR}(v_{i1}, v_{it}) \equiv \rho_v \;\; \forall t = 2, \dots, T \tag{2}$$

(iii)

$$y_{i1} = 1[x'_{i1}\alpha_1 + \cdots + x'_{iT}\alpha_T + \alpha_\delta \delta_i + u_{i1} > 0],$$

$$\text{COR}(u_{i1}, u_{it}) = 0 \;\; \forall t = 2, \dots, T.$$

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 $V(u_{it}) \equiv \sigma_u^2 \; \forall t = 2, \ldots, T$. In (ii) and (iii), $\alpha_1, \ldots, \alpha_T$ 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. A normality assumption has been used for the error terms; for this reason, we will call a model with equations (1) and (2) a 'dynamic probit' from now on.

Turning to the literature review on state dependence, Shaw (1994) uses the Panel Study of Income Dynamics data (PSID) over 1967–87 for the US white females; Shaw uses three separate data (single, single-to-married, and married) and allows different coefficients for four different age groups. Shaw estimates labour participation and work-hour equations separately with lagged work hour, and 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 1984–89 for German males. They use an unrelated-effect maximum likelihood estimator (MLE) under normality assumptions, allowing for

$$\mathrm{COR}(u_{i,t-1}, u_{it}) \neq 0 \quad \text{ with } u_{it} = \xi \cdot u_{i,t-1} + \varepsilon_{it}$$

where ξ is a parameter and ε_{it}s are i.i.d. 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 $1 - y_{it}$ and they let $1 - y_{i,t-1}$ interact with six 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 $V(\delta_i)/V(\varepsilon_{it})$ is 0.023 (insignificant), which means $V(\delta_i)/V(u_{it}) = 0.023/(1 - (-0.32)^2) \simeq 0.026$; i.e. the unobserved heterogeneity is almost non-existent. In a strict sense, this means degeneracy in their MLE.

Hyslop (1999) uses PSID over 1979–85 for married females. Hyslop uses a pseudo-related effect MLE. The linear function for δ_i 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 dealt with as in (ii) above, allowing for

$$\mathrm{COR}(u_{i,t-1}, u_{it}) \neq 0 \quad \text{ with } u_{it} = \xi \cdot u_{i,t-1} + \varepsilon_{it}.$$

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 1, meaning that, as $y_{i,t-1}$ changes from 0 to 1, y_{it}^* changes by one standard deviation (SD) of the error term; a change of 2 SDs would make $y_{it} = 1$ almost certain. Hyslop also finds $V(\eta_i)/\{V(\eta_i) + V(u_{it})\}$ to be about 0.75–0.80. Thus, both state dependence and unobserved heterogeneity matter much. As for ξ, it is estimated to be a small but significant negative number (about -0.22) as in Mühleisen and Zimmermann (1994).

Arulampalam, Booth and Taylor (2000) use the British Household Panel Survey over 1991–95 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 years and 1.41 for men of age 25 years or higher; note that a 1.41 SD change in y^*_{it} because of a change of $y_{i,t-1}$ 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 conclusion.

Phimister, Vera-Toscano and Weersink (2002) use the Canadian Survey of Labour Income and Dynamics over 1993–96 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); i.e. state dependence is strong and the unobserved heterogeneity accounts for 39% of the total error-term variance.

Knight, Harris and Loundes (2002) use the Australian Longitudinal Survey over 1985–88 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, and not the two-stage selection correction approach. β_y values are 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 with these numbers, because interaction terms involving $y_{i,t-1}$ are used); this fact along with significant interaction terms with $y_{i,t-1}$ in Mühleisen and Zimmermann (1994) calls for inclusion of interaction terms involving $y_{i,t-1}$, which is not done adequately in the literature. Secondly, 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 $u_{i2}, \ldots, u_{iT}$ to be i.i.d., which is thus adopted in the remainder of this paper. Thirdly, most empirical studies are for developed countries; hence, it will be interesting to see how the empirical findings for developed countries hold for countries like Korea – an industrial country changing fast while the traditional barriers for women are still strong.

Section II describes DCL and shows that interaction terms with $y_{i,t-1}$ are allowed in DCL, which is not an obvious thing given the difficulty of introducing $y_{i,t-1}$ as a regressor in conditional logit; also shown are the log-likelihood function for dynamic probits and a three-stage algorithm to speed up the convergence of the dynamic probits. Section III describes the data set. Section IV presents the empirical findings. Finally, section V concludes.

II. Methodology

In this section, first, we show DCL in Honoré and Kyriazidou (2000). Secondly, log-likelihood functions for (pseudo-) unrelated-effect dynamics probits are presented. Thirdly, a three-stage algorithm to speed up the dynamic probits is introduced. To be coherent with DCL, 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 change the DCL estimates to be dynamic-probit-comparable by rescaling them properly.

Dynamic conditional logit

Define

$$x_i \equiv (x'_{i1}, x'_{i2}, \ldots, x'_{iT})',$$

and suppose

$$u_{it} \text{ follows logistic distribution independent of } y_{i1}, \delta_i, x_i, \text{ i.i.d. over } i \text{ and } t = 2, \ldots, T \tag{3}$$

u_{i1} does not appear here, because y_{i1} is not specified in DCL. To understand the generality as well as the limitation of DCL, it is necessary to understand the main idea of the estimator in the following.

Let

$$P(y_{i1} = 1 | x_i, \delta_i) \equiv p_{i1},$$

$$P(y_{it} = 1 | x_i, \delta_i, y_{i1}, \ldots, y_{i,t-1}) = \frac{\exp(\beta_y y_{i,t-1} + x'_{it}\beta + \delta_i)}{1 + \exp(\beta_y y_{i,t-1} + x'_{it}\beta + \delta_i)} \quad \text{for} = 2, \ldots, T. \tag{4}$$

Consider two events

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

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

$$P(A|x_i, \delta_i) = p_{i1}^{d_1}(1 - p_{i1})^{1-d_1} \cdot \frac{1}{1 + \exp(\beta_y d_1 + x'_{i2}\beta + \delta_i)} \cdot \frac{\exp(x'_{i3}\beta + \delta_i)}{1 + \exp(x'_{i3}\beta + \delta_i)} \cdot \frac{\exp(d_4(\beta_y + x'_{i4}\beta + \delta_i))}{1 + \exp(\beta_y + x'_{i4}\beta + \delta_i)};$$

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

$$P(B|x_i, \delta_i) = p_{i1}^{d_1}(1 - p_{i1})^{1-d_1} \frac{\exp(\beta_y d_1 + x_{i2}'\beta + \delta_i)}{1 + \exp(\beta_y d_1 + x_{i2}'\beta + \delta_i)} \cdot \frac{1}{1 + \exp(\beta_y + x_{i3}'\beta + \delta_i)} \cdot \frac{\exp(d_4(x_{i4}'\beta + \delta_i))}{1 + \exp(x_{i4}'\beta + \delta_i)}.$$

Define A_0, A_1, B_0, and B_1 such that $P(A|x_i,\delta_i) = A_0/A_1$ and $P(B|x_i,\delta_i) = B_0/B_1$:

$$A_0 \equiv p_{i1}^{d_1}(1 - p_{i1})^{1-d_1} \exp(x_{i3}'\beta + \delta_i)\exp(d_4(\beta_y + x_{i4}'\beta + \delta_i)),$$
$$B_0 \equiv p_{i1}^{d_1}(1 - p_{i1})^{1-d_1} \exp(\beta_y d_1 + x_{i2}'\beta + \delta_i)\exp(d_4(x_{i4}'\beta + \delta_i)),$$
$$A_1 \equiv \{1 + \exp(\beta_y d_1 + x_{i2}'\beta + \delta_i)\}\{1 + \exp(x_{i3}'\beta + \delta_i)\}\{1 + \exp(\beta_y + x_{i4}'\beta + \delta_i)\},$$
$$B_1 \equiv \{1 + \exp(\beta_y d_1 + x_{i2}'\beta + \delta_i)\}\{1 + \exp(\beta_y + x_{i3}'\beta + \delta_i)\}\{1 + \exp(x_{i4}'\beta + \delta_i)\};$$

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 $x_{i3} = x_{i4}$, A_1 and B_1 drop out in the following conditional probabilities that are free of δ_i:

$$\begin{aligned} P(A|x_i, x_{i3} = x_{i4}, \delta_i, A \cup B) &= \frac{A_0}{A_0 + B_0} \\ &= \frac{\exp(x_{i3}'\beta + d_4(\beta_y + x_{i4}'\beta))}{\exp(x_{i3}'\beta + d_4(\beta_y + x_{i4}'\beta)) + \exp(\beta_y d_1 + x_{i2}'\beta + d_4 x_{i4}'\beta)} \\ &= \frac{\exp(x_{i3}'\beta + d_4(\beta_y + x_{i3}'\beta))}{\exp(x_{i3}'\beta + d_4(\beta_y + x_{i3}'\beta)) + \exp(\beta_y d_1 + x_{i2}'\beta + d_4 x_{i3}'\beta)} \\ &= \frac{1}{1 + \exp(\beta_y(d_1 - d_4) + (x_{i2} - x_{i3})'\beta)}; \end{aligned}$$

$$P(B|x_i, x_{i3} = x_{i4}, \delta_i, A \cup B) = \frac{B_0}{A_0 + B_0} = \frac{\exp(\beta_y(d_1 - d_4) + (x_{i2} - x_{i3})'\beta)}{1 + \exp(\beta_y(d_1 - d_4) + (x_{i2} - x_{i3})'\beta)}.$$

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

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

$$\sum_i 1[y_{i2} + y_{i3} = 1]1[x_{i3} = x_{i4}] \ln\left[\frac{\exp\{\beta_y(y_{i1} - y_{i4}) + (x_{i2} - x_{i3})'\beta\}^{y_{i2}}}{1 + \exp\{\beta_y(y_{i1} - y_{i4}) + (x_{i2} - x_{i3})'\beta\}}\right].$$

The estimator is $\sqrt{N}$-consistent and asymptotically normal as other ‘regular’ MLEs.

As δ_i is removed, DCL allows an arbitrary relation between δ_i and x_i. Moreover, *interaction terms involving* $y_{i,t-1}$ *are allowed*. For this, add $y_{i,t-1}z'_{it}\beta_{yz}$ (z_{it} is a vector of regressors and β_{yz} is a parameter vector) to the last two terms of A_1 and B_1 to get

$$\{1+\exp(x'_{i3}\beta+\delta_i)\}\cdot\{1+\exp(\beta_y+z'_{i4}\beta_{yz}+x'_{i4}\beta+\delta_i)\},$$
$$\{1+\exp(\beta_y+z'_{i3}\beta_{yz}+x'_{i3}\beta+\delta_i)\}\cdot\{1+\exp(x'_{i4}\beta+\delta_i)\};$$

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

$$\sum_i 1[y_{i2}+y_{i3}=1]1[x_{i3}=x_{i4}]1[z_{i3}=z_{i4}]$$
$$\cdot\ln\left[\frac{\exp\{\beta_y(y_{i1}-y_{i4})+(y_{i1}z_{i1}-y_{i4}z_{i4})'\beta_{yz}+(x_{i2}-x_{i3})'\beta\}^{y_{i2}}}{1+\exp\{\beta_y(y_{i1}-y_{i4})+(y_{i1}z_{i1}-y_{i4}z_{i4})'\beta_{yz}+(x_{i2}-x_{i3})'\beta\}}\right]. \quad (5)$$

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

The main restriction of DCL other than equation (3) is $x_{i3} = x_{i4}$ (and $z_{i3} = z_{i4}$), which has a number of consequences. First, if x_{it} is continuous, then a non-parametric smoothing should replace $1[x_{i3} = x_{i4}]$; this slows down the convergence rate of the estimator and makes the estimator bandwidth-dependent. Secondly, suppose there is a macroshock changing the intercept. Then DCL is not applicable. For this, let τ_t denote the intercept in the regression function, 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 $x_{i3} = x_{i4}$; to put it differently, time dummies cannot be used as regressors. Thirdly, $x_{i3} = x_{i4}$ disallows regressors such as age or job experience. Despite the restrictions for DCL, however, if the two periods ($t = 3, 4$) have stable macroeconomic conditions so that $\tau_3 \simeq \tau_4$ and if age or job experience effect over 1 year is small, then these restrictions may not be too binding for DCL.

One may try to get around the condition $x_{i3} = x_{i4}$ by grouping variables. For continuous variables such as income, essentially this amounts to a non-parametric smoothing, although this non-parametric aspect is often ignored. For variables such as age (and job experience), this will not work, because equation (5) shows that $x_{i2} \neq x_{i3} = x_{i4}$ is needed, which cannot hold for age.

If age dummies are used such as age20 for being in the twenties or not, $\text{age20}_{i3} = \text{age20}_{i4}$ holds mostly other than when the age crosses the boundary point 20 or 30; thus, $x_{i2} \neq x_{i3} = x_{i4}$ can hold for age20_{it}. This may be criticized for being too 'artificial', although age dummies are popular in practice.

For our empirical analysis with DCL, as age is such an important variable, it has to be accounted for somehow: age dummies will be used only as interacting variables with time-varying variables. For example, $y_{i,t-1}\text{age20}_{it}$ will be used, and equation (5) requires $y_{i1}\text{age20}_{i1} \neq y_{i4}\text{age20}_{i4}$ and $\text{age20}_{i3} = \text{age20}_{i4}$, and not $\text{age20}_{i2} \neq \text{age20}_{i3} = \text{age20}_{i4}$; the former is easier to fulfil, because time-varying y_{i1} and y_{i4} are attached. Moreover, with ed3_{it} being a dummy variable for high school graduation, $\text{age20}_{it}\text{ed3}_{it}$ will be used, for which $\text{age20}_{i2}\text{ed3}_{i2} \neq \text{age20}_{i3}\text{ed3}_{i3} = \text{age20}_{i4}\text{ed3}_{i4}$, not $\text{age20}_{i2} \neq \text{age20}_{i3} = \text{age20}_{i4}$, should hold. Admittedly, this is a somewhat *ad hoc* way of using age_{it} in the model; accounting for age_{it} in DCL is one of the major weaknesses in applying DCL to individual data.

Dynamic probit

For unrelated-effect dynamic probit, the main assumptions are that

1 δ_i is independent of $u_{i1}, \ldots, u_{iT}$, and $(\delta, u_{i1}, \ldots, u_{iT})$ is independent of x_i;
2 $u_{i2}, \ldots, u_{iT}$ are i.i.d. $N(0, \sigma_u^2)$ and independent of u_{i1} that follows $N(0, \sigma_1^2)$;
3 δ_i follows $N(0, \sigma_\delta^2)$.

Define

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

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

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

Dividing the period-1 latent equation by σ_1 and the period-t equation by σ_u, we get the log-likelihood function

$$\sum_i \ln\left[\int \Phi\left\{\left(x_1' \frac{\alpha}{\sigma_1} + \frac{\delta\alpha_\delta}{\sigma_1}\right)(2y_{i1} - 1)\right\} \cdot \prod_{t=2}^{T} \Phi\left\{\left(y_{i,t-1}\frac{\beta_y}{\sigma_u} + y_{i,t-1}z_{it}'\frac{\beta_{yz}}{\sigma_u} + x_{it}'\frac{\beta}{\sigma_u} + \frac{\delta}{\sigma_u}\right)(2y_{it} - 1)\right\}\phi\left(\frac{\delta}{\sigma_\delta}\right)\sigma_\delta^{-1}\text{d}\delta\right] \tag{6}$$

where only x_1 is used in the y_{i1} equation; if desired, x_i can be used without difficulty there. Note that $\phi(\delta/\sigma_\delta)\sigma_\delta^{-1}$ is the density for δ.

To show the identified parameters better, replace δ in the integrand of equation (6) with $(\delta/\sigma_\delta)\sigma_\delta$ to get

$$\sum_i \ln\left[\int \Phi\left\{\left(x_1'\frac{\alpha}{\sigma_1}+\frac{\delta}{\sigma_\delta}\frac{\alpha_\delta\sigma_\delta}{\sigma_1}\right)(2y_{i1}-1)\right\}\right.$$
$$\left.\cdot\prod_{t=2}^{T}\Phi\left\{\left(y_{i,t-1}\frac{\beta_y}{\sigma_u}+y_{i,t-1}z_{it}'\frac{\beta_{yz}}{\sigma_u}+x_{it}'\frac{\beta}{\sigma_u}+\frac{\delta}{\sigma_\delta}\frac{\sigma_\delta}{\sigma_u}\right)(2y_{it}-1)\right\}\phi\left(\frac{\delta}{\sigma_\delta}\right)\sigma_\delta^{-1}\mathrm{d}\delta\right]. \tag{7}$$

As $\zeta \equiv \delta/\sigma_\delta$ is $N(0, 1)$, rewrite equation (7) as

$$\sum_i \ln\left[\int \Phi\left\{\left(x_1'\frac{\alpha}{\sigma_1}+\zeta\frac{\alpha_\delta\sigma_\delta}{\sigma_1}\right)(2y_{i1}-1)\right\}\right.$$
$$\left.\cdot\prod_{t=2}^{T}\Phi\left\{\left(y_{i,t-1}\frac{\beta_y}{\sigma_u}+y_{i,t-1}z_{it}'\frac{\beta_{yz}}{\sigma_u}+x_{it}'\frac{\beta}{\sigma_u}+\zeta\frac{\sigma_\delta}{\sigma_u}\right)(2y_{it}-1)\right\}\phi(\zeta)\mathrm{d}\zeta\right] \tag{8}$$

which is the log-likelihood function to be used for unrelated-effect dynamic probit. The identified parameters here are

$$\frac{\alpha}{\sigma_1},\ \frac{\alpha_\delta\sigma_\delta}{\sigma_1},\ \frac{\beta_y}{\sigma_u},\ \frac{\beta_{yz}}{\sigma_u},\ \frac{\beta}{\sigma_u},\ \frac{\sigma_\delta}{\sigma_u}; \tag{9}$$

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

For pseudo-related-effect dynamic probit, we can use

$$\delta_i = x_{i1}'\mu_1 + \cdots + x_{iT}'\mu_T + \eta_i,$$

or a simpler version

$$\delta_i = \bar{x}_i'\mu + \eta_i, \quad \bar{x}_i \equiv T^{-1}\sum_{t=1}^{T} x_{it}.$$

As the former is computationally too demanding, and although it can be dealt with in principle as in Lee (2002), we adopt the simpler version that includes the restriction $\mu_1 = \cdots = \mu_T \equiv \mu_0$:

$$\delta_i = \left(\sum_t x_{it}'\right)\mu_0 + \eta_i = \bar{x}_i'(\mu_0 T) + \eta_i = \bar{x}_i'\mu + \eta_i, \qquad \text{where } \mu \equiv \mu_0 T. \tag{10}$$

Writing x_{it} as $x_{it} - \bar{x}_i + \bar{x}_i$, the effect of x_{it} can be decomposed into two: the temporary (or transitory) effect from $x_{it} - \bar{x}_i$ and the permanent effect from $\bar{x}_i$; $\mu = \mu_0 T$ implies that the permanent effect is the sum of 'one-shot time-invariant' effects over T periods. The permanent effect is a level change, for which expressions such as 'tendency' or 'propensity' are often used.

Substitute $\delta_i = \bar{x}_i'\mu + \eta_i$ in equation (6) to get

$$\sum_i \ln\left[\int \Phi\left\{\left(x_1'\frac{\alpha}{\sigma_1} + (\bar{x}_i'\mu + \eta_i)\frac{\alpha_\delta}{\sigma_1}\right)(2y_{i1}-1)\right\}\cdot\prod_{t=2}^{T}\Phi\left\{\left(y_{i,t-1}\frac{\beta_y}{\sigma_u}\right.\right.\right.$$
$$\left.\left.\left. + y_{i,t-1}z_{it}'\frac{\beta_{yz}}{\sigma_u} + x_{it}'\frac{\beta}{\sigma_u} + (\bar{x}_i'\mu + \eta_i)\frac{1}{\sigma_u}\right)(2y_{it}-1)\right\}\phi\left(\frac{\eta}{\sigma_\eta}\right)\sigma_\eta^{-1}\mathrm{d}\eta\right].$$

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

$$\sum_i \ln\left[\int \Phi\left\{\left(x_1'\frac{\alpha}{\sigma_1} + \bar{x}_i'\frac{\mu\alpha_\delta}{\sigma_1} + \zeta\frac{\alpha_\delta\sigma_\eta}{\sigma_1}\right)(2y_{i1}-1)\right\}\right.$$
$$\left.\cdot\prod_{t=2}^{T}\Phi\left\{\left(y_{i,t-1}\frac{\beta_y}{\sigma_u} + y_{i,t-1}z_{it}'\frac{\beta_{yz}}{\sigma_u} + x_{it}'\frac{\beta}{\sigma_u} + \bar{x}_i'\frac{\mu}{\sigma_u} + \zeta\frac{\sigma_\eta}{\sigma_u}\right)(2y_{it}-1)\right\}\phi(\zeta)d\zeta\right], \tag{11}$$

which is the log-likelihood function for pseudo-related-effect dynamic probit.

The identified parameters to be compared with equation (9) are

$$\frac{\alpha}{\sigma_1}, \underline{\frac{\mu\alpha_\delta}{\sigma_1}}, \boxed{\frac{\alpha_\delta\sigma_\eta}{\sigma_1}}, \frac{\beta_y}{\sigma_u}, \frac{\beta_{yz}}{\sigma_u}, \frac{\beta}{\sigma_u}, \underline{\frac{\mu}{\sigma_u}}, \boxed{\frac{\sigma_\eta}{\sigma_u}}; \tag{12}$$

the two underlined terms did not appear in equation (9) and the two boxed terms appeared in equation (9) with σ_η replaced by σ_δ. 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 get to be used twice as regressors in x_{it} and $\bar{x}_i$ for the same equation. Thus the coefficients for the time-constant variables in x_{it} are in fact the sum of the coefficients for x_{it} 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 equation (8) or in (11). The fixed y_{i1} assumption is often used in practice; comparing the estimates under fixed y_{i1} with those under random y_{i1} will show how sensitive the dynamic probit is to the initial period assumption.

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

(i) DCL: equation (5).
(ii) unrelated-effect dynamic probit with y_{i1} fixed: (8) with period 1 removed.
(iii) unrelated-effect dynamic probit with y_{i1} random: (8).
(iv) pseudo-related-effect dynamic probit with y_{i1} fixed: (11) with period 1 removed.
(v) pseudo-related-effect dynamic probit with y_{i1} random: (11).

Integration for dynamic probit can be done numerically. But this turned out to be too time consuming. Instead, we use simulated MLE 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 MLE calls for the random draw number to be infinite. With the number 10, the dynamic probit programs took about 2 hours on a Dell Pentium IV PC; with 35, it took about a day. Hence, with 35, we use the following three-stage algorithm to speed up the program. Numerical integration requires splitting $(-\infty, \infty)$ to, say, 100 intervals, which is more or less the same as using 100 random draws in simulated MLE in terms of computation time.

Three-stage algorithm to speed up dynamic probit

The integration in equation (11) can be done either numerically or with a Monte Carlo simulation. Either way, maximizing equation (11) 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 is helpful.

The idea goes as follows. First, estimate the initial period equation with probit; no lagged response here. Secondly, estimate the remaining parameters while keeping the first-stage parameters intact. Thirdly, 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.

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

$$\sigma_a^2 \equiv V(\alpha_\delta\eta + u_1) = (\alpha_\delta\sigma_\eta)^2 + \sigma_1^2 \Longrightarrow (\alpha_\delta\sigma_\eta/\sigma_a)^2 = 1 - (\sigma_1/\sigma_a)^2. \quad (13)$$

In the first-stage probit, as the error term is $\alpha_\delta\eta + u_1$ when δ is not conditioned on, σ_a is the right normalizing scale factor, differently from σ_1 appearing in equation (11) where δ is conditioned on before finally integrated out. The first-stage probit yields $\widehat{\alpha/\sigma_a}$ and $\widehat{\mu\alpha_\delta/\sigma_a}$.

In the second stage, we need to maximize equation (11) with the first-stage estimates plugged in. For this, rewrite the initial period likelihood contribution in equation (11) as

$$\Phi\left\{\left(x_1'\frac{\alpha}{\sigma_a}\frac{\sigma_a}{\sigma_1} + \bar{x}_i'\frac{\mu\alpha_\delta}{\sigma_a}\frac{\sigma_a}{\sigma_1} + \zeta\frac{\alpha_\delta\sigma_\eta}{\sigma_a}\frac{\sigma_a}{\sigma_1}\right)\cdot(2y_{i1}-1)\right\}$$

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

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

$$\frac{\sigma_a}{\sigma_1}, \frac{\beta_y}{\sigma_u}, \frac{\beta_{yz}}{\sigma_u}, \frac{\beta}{\sigma_u}, \frac{\mu}{\sigma_u}, \frac{\sigma_\eta}{\sigma_u}. \tag{14}$$

In the third stage, combine the first-stage probit and the second-stage estimates to get equation (12) and take one Newton–Raphson step from the estimates; better yet, the third stage can be iterated fully as mentioned above.

III. Data and descriptive statistics

Our data set consists of the first four waves (1998–2001) of the Korea Labour and Income Panel Study (KLIPS). Excluding the dropouts, the data set is balanced with $N = 3{,}882$ for women of age 15–65 years. Although panel data do not have to be balanced for any of our methods so long as the unbalanced panel data do not suffer from sample selection problems, we use the balanced panel data because there are only four waves available that are the minimum number of waves necessary for DCL. If there were more than four-waves and the panel data were unbalanced, then we could use all observed four-wave combinations for DCL (see Honoré and Kyriazidou, 2000, p. 852). For instance, if a person is observed at $t = 1, 2, 3, 4, 6$, then this person's observations can be used multiple times, say, (1, 2, 3, 4), (1, 2, 3, 6), (2, 3, 4, 6) and so on, so long as the events in the indicator functions of equation (5) hold for each combination.

Other than the response variable for working or not, the following regressors are used: age, the number of children aged 1–3 (ch1), 4–7 (ch2), and 8–13 years (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 ($= \text{age}^2/100$), age20 for being in the twenties or not, age30 for being in the thirties or not, and age40 for being in the forties or above are also used as regressors. ed1 is the base education case and thus not used in estimation.

Table 1 shows the descriptive statistics over the 4 years. The labour participation rate is relatively low (46%). The high-school graduation is the majority. Income in Table 1 is not in log, but in Korean-Won $\times$ 10,000, which is roughly \$8; thus, the median household income other than the female's own income is \$8,872 $= 1{,}109 \times 8$. About 75% are married.

TABLE 1

Descriptive statistics

Variable	*Mean*	*SD*	*Min.*	*Med.*	*Max.*
y	0.457	0.498	0	0	1
Age	37.51	12.87	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	1,109	12,000,000
jtr	0.036	0.187	0	0	1
mar	0.752	0.432	0	1	1

TABLE 2

Change in labour participation

Years	*0 → 0*	*0 → 1*	*1 → 0*	*1 → 1*	*Sum*
1998–99	46	14	6	34	100
1999–2000	45	7	9	39	100
2000–2001	45	9	6	40	100

Before we examine Table 2, ignore all regressors for a while and consider

$$y_{it} = 1[\beta_y y_{i,t-1} + v_{it} \geq 0], \quad v_{it} \sim N(0,1).$$

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 a woman starts to work (maybe because $v_{i,t-1}$ takes a big positive value), the stronger the state dependence is, the less likely she goes back to non-working.

It is also helpful to examine, for parameters β_{y0} and β_{y1},

$$y_{it} = 1[\beta_{y0}(1-y_{i,t-1}) + \beta_{y1}y_{i,t-1} + v_{it} \geq 0] = 1[\beta_{y0} + (\beta_{y1} - \beta_{y0})y_{i,t-1} + v_{it} \geq 0].$$

That is, $\beta_y = \beta_{y1} - \beta_{y0}$ is in fact a combination of the two state dependence parameters for working and non-working; a negative β_{y0} and a positive β_{y1}

will make β_y greater. Now 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 ($x_{it}'\beta > 0$) favourable position to work; other than this, an analogous interpretation can be given to β_y.

Table 2 shows that the percentage of labour participation changes over the 4-year period; e.g. during 1998–99, the proportion of $0 \rightarrow 1$ in $0 \rightarrow 0$ or $0 \rightarrow 1$ is $14/60 = 0.23$ whereas the proportion of $1 \rightarrow 1$ in $1 \rightarrow 0$ or $1 \rightarrow 1$ is $34/40 = 0.85$. This suggests a strong state dependence, but without x_{it} controlled for, nothing definite can be said.

IV. Empirical findings for Korean female labour participation

Table 3 shows the estimation results for DCL. We will base our interpretation on the '1.8-normalized' estimates in the last column. Before we proceed further, two remarks are in order. First, although we tried to include as many variables to be used for DCL as possible in Table 3, the temporal variation in some variables was too small for the coefficients to be estimable; hence, those variables are excluded from Table 3. Secondly, income is not used for DCL, because it is a continuous variable requiring smoothing in DCL; 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 3, we tried many interaction terms with y_{t-1} and only ed4 came out significant; mar, age20, age30, and age40 are included for later

TABLE 3

Dynamic conditional logit

	Estimate	*t-value*	*Estimate/1.8*
y_{t-1}	1.077	(1.30)	0.598
y_{t-1} * mar	0.209	(0.37)	0.116
y_{t-1} * ed4	2.721	(2.51)	1.512
y_{t-1} * age20	−0.832	(−0.93)	−0.462
y_{t-1} * age30	−0.814	(−0.82)	−0.452
y_{t-1} * age40	0.363	(0.35)	0.202
ch1	0.741	(1.15)	0.412
ch2	−1.029	(−1.66)	−0.572
ch3	−0.375	(−0.82)	−0.208
ed3	0.122	(0.14)	0.068
ed5	0.168	(0.14)	0.093
age20 * ed3	−0.921	(−1.49)	−0.512
age30 * ed3	−1.031	(−1.17)	−0.573
cert	−0.012	(−0.01)	−0.007
jtr	1.110	(1.62)	0.617
Conditional log-like		−226.954	

comparisons with dynamic probits. Other than for ed4 = 1, the state dependence is 0.598; 0.598 (with the t-value 1.30) falls close to the low end of the state dependence numbers seen in the literature for developed countries. For women with a junior college degree (ed4 = 1), the state dependence is 2.11 = 0.598 + 1.512, which indicates almost no change in work status. This may be because of the fact that the education at junior college tends to be job-oriented and job-specific. For women in their twenties, the state dependence is quite small: 0.136 = 0.598 − 0.462. The state dependence for women in their thirties is also small: 0.146 = 0.598 − 0.452. Other than state dependence, ch2 and jtr have fairly significant coefficients, comparable with that of y_{t-1} in magnitude: ch2 decreases labour participation whereas job training increases it as one would expect. Only about 5% of the women in our data change ed3 (from 0 to 1) and about 1% change ed5, and probably as a consequence, ed3 and ed5 have insignificant estimates.

Table 4 presents unrelated-effect dynamic probit results. The columns with 'y_1-fixed' are vastly different from those 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. Thus we will base our interpretation of unrelated-effect dynamic probit on the last two columns; the estimates for the initial period are of the 'reduced-form type' and thus difficult to interpret.

The magnitude of y_{t-1} in Table 4 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 almost the same as those for DCL: state dependence is almost zero for single women in their twenties and thirties. Age has the usual up and down pattern for labour 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.

The effects of 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 labour participation is

$$\text{For women in their twenties: } 1.390 = 2.227 - 0.837$$
$$\text{For women in their thirties: } 0.318 = 1.155 - 0.837$$
$$\text{For women in their forties or above: } -0.837.$$

This implies that, for mature women, college education is a hindrance for labour participation, perhaps because they are less willing to take easily available low-paying jobs.

TABLE 4

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	−4,966.47	−7,080.355	

The estimate for cert is significant but small. If income increases by 1%, then the labour participation propensity decreases by 0.031 ∗ SD(error), which is quite small. Job training has a significant estimate (0.516) the magnitude of which is close to that for DCL (0.617). Marriage with the estimate −1.327 seems to be the most detrimental to female labour participation. The SD ratio of the time-invariant error to time-variant error is about 1.344; i.e. the ratio of time-invariant error variance to the total error variance is $0.643 = 1.344^2/(1.344^2 + 1)$, which is close to the estimates in the literature.

Table 5 presents pseudo-related-effect dynamic probit results. As in Table 4, the columns with 'y_1-fixed' are vastly different from the columns with 'y_1-random': omitting the initial period equation introduces huge biases

TABLE 5

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)	$\overline{\text{ch1}}$ −1.289 (−6.54)
ch2	1.110 (4.69)	0.045 (0.35)	−0.002 (−0.03)	$\overline{\text{ch2}}$ −0.286 (−1.68)
ch3	1.279 (7.68)	0.165 (1.52)	−0.135 (−1.67)	$\overline{\text{ch3}}$ 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)	$\overline{\text{cert}}$ −0.089 (−0.45)
inc	0.106 (1.35)	−0.079 (−4.65)	−0.025 (−3.75)	$\overline{\text{inc}}$ −0.204 (−7.37)
jtr	0.703 (4.41)	0.051 (0.28)	0.298 (2.40)	$\overline{\text{jtr}}$ 2.030 (6.16)
mar	0.362 (1.47)	−1.391 (−3.60)	−1.485 (−4.46)	$\overline{\text{mar}}$ 0.656 (2.07)
σ_η/σ_u	−0.039 (−0.37)	1.342 (17.53)	1.395 (15.42)	
Log-likelihood	−4,915.768		−6,988.534	

again. As in Table 4, we will base our interpretation on the columns 'y_1-random'. The last two columns of Table 5 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 estimates for y_{t-1} and its interaction terms are little different from those in Table 4, and the comments made for Table 4 apply to Table 5 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 labour participation.

The estimate for ch1 is almost zero whereas $\overline{\text{ch1}}$ has a significant negative coefficient that is also large in magnitude (more than twice the effect of y_{t-1}). This means that the temporary effect of having an infant on y_t is almost zero but the permanent effect is strong. But what can be the permanent component of ch1? If T is large, $\overline{\text{ch1}}$ has to be very small; i.e., the permanent component cannot be the persistent level of ch1, which is almost zero in the long run. Rather, the permanent component – the 'tendency' to have infants – could be the lack of job-related skills, which affects $\overline{\text{ch1}}$ positively and y_t negatively. Recalling equation (10), the 'permanent infant effect' per period is obtained as $-1.289/4 = -0.32$ with $T = 4$.

If the permanent income $\overline{\text{inc}}$ increases by 1%, then this decreases labour participation propensity by 0.204 $*$ SD(error). The permanent income effect per period is $-0.204/4 = -0.051$ that is twice the temporary effect -0.025. In Table 4, the income effect is -0.031, which may be -0.025 in Table 5 plus the bias term from the omitted $\overline{\text{inc}}$. In fact, a similar statement can be made for other variables as well: e.g. ch1's estimate in Table 4 is -0.238, which may be 0.030 in Table 5 plus the bias from the omitted $\overline{\text{ch1}}$.

The permanent component $\overline{\text{jtr}}$ of job training has a significant estimate 2.03; the magnitude is more than three times the estimate for y_{t-1} (0.571). The temporary effect 0.298 is significant but about seven times smaller in magnitude than the permanent effect. As in $\overline{\text{ch1}}$, the permanent-level $\overline{\text{jtr}}$ should go to zero as T gets larger; $\overline{\text{jtr}}$ had better be viewed as eagerness to work that is a time-constant unobserved trait, whereas jtr is for the effect of taking a job-training newly. The 'permanent job-training effect' per period is $2.03/4 = 0.51$, which is greater than the temporary effect 0.298. The DCL estimate 0.617 for jtr is much bigger than 0.298 in Table 5, suggesting that controlling for the eagerness to work with $\overline{\text{jtr}}$ may be inadequate.

Interestingly, the long-term marriage effect from $\overline{\text{mar}}$ (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 labour participation. The 'permanent marriage effect' per period is $0.656/4 = 0.164$, the magnitude of which is about half the permanent infant effect per period.

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:

$$\delta_i = x'_{i1}\mu_1 + \cdots + x'_{iT}\mu_T + \eta_i.$$

This shows that, part of u_{it} related to $x_{i1}, \ldots, x_{iT}$ may have been taken out as well by $\bar{x}_i$; i.e. σ_u can also decrease in the pseudo-related-effect dynamic probit.

Finally, Table 6 presents the three-stage estimation results. As 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

TABLE 6

Pseudo-related-effect dynamic probit (threestages)

	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)	$\overline{\text{ch1}}$ −1.449 (−6.67)
ch2	0.037 (0.29)	−0.006 (−0.07)	$\overline{\text{ch2}}$ −0.327 (−1.88)
ch3	0.153 (1.41)	−0.126 (−1.56)	$\overline{\text{ch3}}$ 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)	$\overline{\text{cert}}$ −0.095 (−0.49)
inc	−0.079 (−4.69)	−0.025 (−3.84)	$\overline{\text{inc}}$ −0.191 (−7.23)
jtr	0.009 (0.05)	0.297 (2.42)	$\overline{\text{jtr}}$ 1.914 (6.30)
mar	−1.139 (−3.49)	−1.440 (−4.55)	$\overline{\text{mar}}$ 0.722 (2.33)
σ_η/σ_u	1.291 (17.79)	1.416 (16.02)	
Log-likelihood	−6,986.907		

in the last stage, we did a full iteration. This 'finer' estimation effort resulted in the log-likelihood function value increasing slightly from −6,988.534 in Table 5 to −6,986.907 in Table 6. Overall, Tables 5 and 6 are not much different, and the comments made for Table 5 apply to Table 6 as well; the only exception seems to be y_{t-1}ed4, which is now significant with a slightly bigger estimate.

V. Conclusions

We estimated dynamic labour participation models for Korean women, using dynamic conditional logit, and unrelated or pseudo-related dynamic probit. On the methodological front, we showed that interaction terms with y_{t-1} are allowed in dynamic conditional logit and suggested a three-stage estimation procedure to reduce the computation time for dynamic probits. We also showed that the probit estimators treating the initial period response as fixed are highly biased and should not be used.

Despite the differences in the models, we obtained more or less coherent results across the models. Lagged response y_{t-1} matters much with unobserved heterogeneity allowed for. Moreover, y_{t-1} interacts with some regressors: marriage dummy, junior (technical) college dummy, and age group dummies for women in their twenties and thirties. The state dependence is about 0.6, and it is 0.2 to 0.7 higher for the married and about 0.5 to 0.6 smaller for women in their twenties or thirties; thus, there is 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 labour participation, but college education can be a hindrance for women in their forties or higher. The transitory effect of getting married is highly negative (-1.4) whereas the long-term effect of being married is positive (about 0.16 per period). These findings lead to policy implications: junior college education more geared for job skills be supported, and policy efforts to increase female labour participation be directed at women of relatively high age.

References

Arulampalam, W., Booth, A. L. and Taylor, M. P. (2000). 'Unemployment persistence', *Oxford Economic Papers*, Vol. 52, pp. 24–50.

Chamberlain, G. (1984). 'Panel data', in Griliches Z. and Intrilligator M. (eds), *Handbook of Econometrics*, Vol. 2, North Holland, Amsterdam, pp. 1247–1318.

Honoré, B. E. and Kyriazidou, E. (2000). 'Panel data discrete choice models with lagged dependent variables', *Econometrica*, Vol. 68, pp. 839–874.

Hyslop, D. R. (1999). 'State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women', *Econometrica*, Vol. 67, pp. 1255–1294.

Knight, S., Harris, M. and Loundes, J. (2002). 'Dynamic relationships in the Australian labour market: heterogeneity and state dependence', *Economic Record* Vol. 78, pp. 284–298.

Lee, M. J. (2002). *Panel Data Econometrics: Method-of-moments and Limited Dependent Variables*, Academic Press, San Diego, CA.

Mühleisen, M. and Zimmermann, K. F. (1994). 'A panel analysis of job changes and unemployment', *European Economic Review*, Vol. 38, pp. 793–801.

Phimister, E., Vera-Toscano, E. and Weersink, A. (2002). 'Female participation and labor market attachment in rural Canada', *American Journal of Agricultural Economics*, Vol. 84, pp. 210–221.

Shaw, K. (1994). 'The persistence of female labor supply', *Journal of Human Resources*, Vol. 29, pp. 348–378.