# DCDP Models as a Framework for Policy Evaluation

Petra E. Todd[1]

[1]University of Pennsylvania

January, 2013

# Introduction

- Based on *Handbook of Labor Economics Chapter* by Michael Keane, Petra Todd and Kenneth Wolpin
- Introduction to the methods of structural estimation of discrete choice dynamic programming models (DCDP) for policy evaluation purposes.
- Surveys the contributions of applications of these methods in labor economics.
- The development of DCDP estimation methods over the last 25 years opened up new frontiers for empirical research in labor economics, industrial organization, economic demography, health economics, development economics and political economy

# Early papers

- Gotz and McCall F(1984) considered the decision to re-enlist in the military
- Miller (1984) the decision to change occupations
- Pakes (1986) the decision to renew a patent
- Rust (1987) the decision to replace a bus engine and
- Wolpin (1984) the decision to have a child

# Introduce to the solution and estimation of DCDP models

- The development of the DCDP empirical framework was a straightforward and natural extension of the static discrete choice empirical framework.
- The latent variable specification is the building block for all economic models of discrete choice.
- Consider a binary choice model in which an economic agent with imperfect foresight, denoted by $i$, makes a choice at each discrete period $t$, from $t = 1, .. T$, between two alternatives $d_{it} \in \{0, 1\}$.
- Might be the choice of whether to accept a job offer or remain unemployed or whether to attend college or enter the labor force or whether to participate in a training program.

- The outcome is determined by whether a latent variable, $v_{it}^*$, reflecting the difference in the payoffs of the $d_{it} = 1$ and $d_{it} = 0$ alternatives, crosses a scalar threshold value, WLOG is taken to be zero.

- The preferred alternative is the one with the largest payoff, i.e., where $d_{it} = 1$ if $v_{it}^* \geq 0$ and $d_{it} = 0$ otherwise.

The latent variable may be a function of three types of variables:

- $\widetilde{D}_{it}$ is a vector of the history of past choices $(d_{i\tau} : \tau = 1, ..., t-1)$
- $\widetilde{X}_{it}$ is a vector of contemporaneous and lagged values of $J$ additional variables $(X_{ij\tau} : j = 1, ..., J;\ \tau = 1, ..., t)$ that are not chosen by the agent and that enter the decision problem
- $\widetilde{\varepsilon}_{it}$ $(\varepsilon_{i\tau}\text{: } \tau = 1, ...t)$ is a vector of contemporaneous and lagged unobservables that also enter the decision problem.

The agent's decision rule is:

$$\begin{aligned} d_{it} &= 1 \; if \quad v^*_{it}(\widetilde{D}_{it}, \widetilde{X}_{it}, \widetilde{\varepsilon}_{it}) \geq 0, \\ &= 0 \; if \quad v^*_{it}(\widetilde{D}_{it}, \widetilde{X}_{it}, \widetilde{\varepsilon}_{it}) < 0. \end{aligned} \tag{1}$$

-All empirical binary choice models, dynamic or static, are special cases of this formulation.

- The underlying behavioral model that generated the latent variable is dynamic if agents are forward looking, e.g. either $v^*_{it}$ contains past choices, $\widetilde{D}_{it}$, or unobservbles, $\widetilde{\varepsilon}_{it}$, that are serially correlated.

Researchers may have a number of different goals, such as:

- Test a prediction of the theory, such as how an observable variable in $v_{it}^*$ affects $d_{it}$.
- Determine the impact of a change in $\widetilde{D}_{it}$ or $\widetilde{X}_{it}$ on choices.
- Determine the impact of a change in something not in $\widetilde{D}_{it}$ or $\widetilde{X}_{it}$ on choices.

# Example: Labor force participation of married women

Consider the following static model of the labor force participation decision of a married woman.

Assume a unitary model. The couple's utility is

$$U_{it} = U(c_{it}, 1 - d_{it}; \kappa_{it}(1 - d_{it}), \varepsilon_{it}(1 - d_{it})), \tag{2}$$

$c_{it}$ is household $i$'s consumption at period $t$,
$d_{it} = 1$ if the wife works, $= 0$ otherwise
$\kappa_{it}$ are observables affecting utility from leisure
$\varepsilon_{it}$ (serially uncorrelated) unobservable factors that affect the couple's valuation of the wife's leisure (or home production).

$\kappa_{it}$ might include, among other things, the number of young children in the household, $n_{it}$ and the duration of marriage, $t$.

The preference unobservable is assumed to be randomly drawn independently over time.

The utility function has the usual properties:
$\partial U/\partial C > 0, \partial^2 U/\partial C^2 < 0, U(C,1) > U(C,0)$.

- The wife receives a wage offer of $w_{it}$ in each period $t$

- If the wife works, the household incurs a per-child child-care cost, $\pi$, which is assumed to be time-invariant, unobserved, and the same for all households.

The husband is assumed to work each period and to generate income $y_{it}$.

- The budget constraint is

$$c_{it} = y_{it} + w_{it} d_{it} - \pi n_{it} d_{it}. \tag{3}$$

Wage offers are not observed for non-workers, so specify a wage offer function:

$$w_{it} = w(z_{it}, \eta_{it}), \tag{4}$$

- $z_{it}$ would typically contain educational attainment and "potential" work experience (age - education - 6).

- $\eta_{it}$,the wage shock is assumed to be randomly drawn independently over time.

- Unobservable factors that enter the couple's utility function, $(\varepsilon_{it})$, and $(\eta_{it})$ are assumed to have a joint distribution $F$.

Substituting (3) into (2) using (4) yields

$$U_{it} = U(y_{it} + w(z_{it}, \eta_{it})d_{it} - \pi n_{it} d_{it}, 1 - d_{it}; \kappa_{it}(1 - d_{it}), \varepsilon_{it}(1 - d_{it})), \tag{5}$$

from which we get alternative-specific utilities, $U_{it}^1$ if the wife works and $U_{it}^0$ if she does not, namely

$$\begin{aligned} U_{it}^1 &= U(y_{it} + w(z_{it}, \eta_{it}) - \pi n_{it}, 0), \\ U_{it}^0 &= U(y_{it}, 1; \kappa_{it}, \varepsilon_{it}). \end{aligned} \tag{6}$$

The latent variable function, the difference in utilities, $U_{it}^1 - U_{it}^0$, is

$$v_{it}^* = v^*(y_{it,} z_{it}, n_{it}, \kappa_{it}, \varepsilon_{it}, \eta_{it}) \tag{7}$$

The participation decision is determined by the sign of the latent variable:
$d_{it} = 1$ if $v_{it}^* \geq 0, d_{it} = 0$ otherwise.

- the household's state space, $\Omega_{it}$, consists of all of the determinants of the household's decision, that is, $y_{it}, z_{it}, n_{it}, \kappa_{it}, \varepsilon_{it}, \eta_{it}$.

- the part of the state space observable to the researcher, $\Omega_{it}^{-}$, consists of $y_{it}, z_{it}, n_{it}, \kappa_{it}$.

- Define $S(\Omega_{it}^{-}) = \{\varepsilon_{it}, \eta_{it} : v^{*}(\varepsilon_{it}, \eta_{it}; \Omega_{it}^{-}) > 0\}$ to be the set of $\{\varepsilon_{it}, \eta_{it}\}$ that induces a couple with a given observable state space $(\Omega_{it}^{-})$ to choose $d_{it} = 1$.

- Assuming that the elements of $\Omega_{it}^{-}$ are distributed independently of $\varepsilon_{it}$ and $\eta_{it}$, the probability of choosing $d_{it}=1$, conditional on $\Omega_{it}^{-}$, is:

$$\Pr(d_{it}=1|\Omega_{it}^{-}) = \int_{S(\Omega_{it}^{-})} dF = G(y_{it}, z_{it}, n_{it}, \kappa_{it}), \qquad (8)$$

where $\Pr(d_{it}=0|\Omega_{it}^{-}) = 1 - \Pr(d_{it}=1|\Omega_{it}^{-})$.

- $G(y_{it}, z_{it}, n_{it}, \kappa_{it})$ is a composite of three elements of the model, which comprise the *structure*: $U(\cdot), w(\cdot), F$.

- Structural estimation (S) is concerned with recovering some or all of the structural elements of the model.

- Non-structural (NS) estimation is concerned with recovering $G(\cdot)$.

- Each of these estimation approaches can adopt auxiliary assumptions in terms of parametric (P) forms for some or all of the structural elements or for $G(\cdot)$ or be non-parametric (NP).

Thus, there are four possible approaches to estimation:

- nonparametric/nonstructural (NP-NS)
- parametric/nonstructural (P-NS)
- nonparametric/structural (NP-S)
- parametric/nonstructural (P-S).

Consider how each of these approaches can be used to address the following goals:

Goal 1. Test the model by testing whether the probability of working is increasing in the wage offer.

Goal 2. Determine the impact of changing any of the state variables in the model on the participation probability (requires taking derivative)

Goal 3. Determine the effect on the participation probability of varying something that does not vary in the data, e.g. the effect of a child care subsidy.

# Non-Parametric, Non-Structural:

We can estimate $G(y_{it}, z_{it}, n_{it}, \kappa_{it})$ non-parametrically.

Goal 1: We need to be able to vary the wage offer independently of other variables that affect participation. To do that, there must be an exclusion restriction, a variable in $z_{it}$ that is not in $\kappa_{it}$.

- If we observed wage offers for everyone, the test of the prediction could be performed directly without an exclusion restriction.

Goal 2. Given a nonparametric estimate of $G$, we can determine the effect on participation of a change in any of the variables *within the range of the data*.

Goal 3: It is not possible to separately identify $G$ and $\pi$. To see that note that because it is $\pi n$ that enters $G$, ($G_n = \pi G_{\pi n}$), knowledge of $G_n$ does not allow one to separately identify $G_{\pi n}$ and $\pi$. We thus cannot perform the child care subsidy policy experiment.

# Parametric, Non-Structural:

Choose a functional form for $G$, (bounded between 0 and 1), e.g., a cumulative standard normal in which the variables in $\Omega_{it}^{-}$ enter as a single index.

Goal 1: Because of partial observability of wage offers, testing the model's prediction requires an exclusion a variable in $z_{it}$ that is not in $\kappa_{it}$.

Goal 2. It is possible, given an estimate of $G$, to determine the effect on participation of a change in any of the variables not only within but also outside the range of the data.

Goal 3: It is not possible to separately identify $\pi$ from variation in $n_{it}$ because $\pi n_{it}$ enters $G$.

# Non-Parametric, Structural

In this approach, one would typically attempt to separately identify $U(\cdot), w(\cdot), F$ from (8) without imposing auxiliary assumption about those functions.

This is clearly infeasible. Even if we had data on all wage offers so that $w(\cdot)$ and the marginal distribution, $F_\eta$, were non-parametrically identified, we could still not identify $U(\cdot)$ and $F$. Maybe possible to do policy evaluation imposing structural model

but without fully recovering structure (will discuss later).

# Parametric, Structural:

Consider the case in which all of the structural elements are parametric, specified as:

$$U_{it} = c_{it} + \alpha_{it}(1 - d_{it}) \text{ with } \alpha_{it} = \kappa_{it}\beta + \varepsilon_{it}, \tag{9}$$

$$c_{it} = y_{it} + w_{it}d_{it} - \pi n_{it}d_{it}, \tag{10}$$

$$w_{it} = z_{it}\gamma + \eta_{it}, \tag{11}$$

$$f(\varepsilon_{it}, \eta_{it})\tilde{\ } N(0, \Lambda), \tag{12}$$

where $\Lambda = \begin{pmatrix} \sigma_\varepsilon^2 & \cdot \\ \sigma_{\varepsilon\eta} & \sigma_\eta^2 \end{pmatrix}$.

The difference in utilities, $U_{it}^1 - U_{it}^0$, is

$$\begin{aligned} v_{it}^*(z_{it}, n_{it}, \kappa_{it}, \eta_{it}, \varepsilon_{it}) &= z_{it}\gamma - \pi n_{it} - \kappa_{it}\beta + \eta_{it} - \varepsilon_{it} \\ &= \xi_{it}^*(\Omega_{it}^-) + \xi_{it}, \end{aligned} \tag{13}$$

where $\xi_{it} = \eta_{it} - \varepsilon_{it}$, $\xi_{it}^*(\Omega_{it}^-) = z_{it}\gamma - \pi n_{it} - \kappa_{it}\beta$ and $\Omega_{it}^-$ now consists of $z_{it}$, $n_{it}$ and $\kappa_{it}$.

-Additive error not required (would break down in wages are in log form or utility not linear in consumption).

- The linearity and separability of consumption in the utility function implies that husband's income does not enter $v_{it}^*$ and, thus, does not affect the participation decision.

The likelihood function, incorporating the wage information for those women who work, is

$$L(\theta; \kappa_{it_i}, z_{it_i}, \pi_i n_{it_i}) = \tag{14}$$

$$\prod_{i=1}^{I} \Pr(d_{it_i} = 1, w_{it_i} |\ \Omega_{it}^{-})^{d_{it_i}} \Pr(d_{it_i} = 0 | \Omega_{it_i}^{-})^{1-d_{it_i}} = \tag{15}$$

$$\prod_{i=1}^{I} \Pr(\ \xi_{it_i} \geq -\xi_{it_i}^{*}(\Omega_{it_i}^{-}), \eta_{it_i} = w_{it_i} - z_{it_i}\gamma)^{d_{it_i}} \times \tag{16}$$

$$\Pr(\xi_{it} < -\xi_{it}^{*}(\Omega_{it}^{-}))^{1-d_{it_i}}.$$

The parameters to be estimated include $\beta, \gamma, \pi, \sigma_\varepsilon^2, \sigma_\eta^2$, and $\sigma_{\varepsilon\eta}$

- Not possible to separately identify the child care cost $\pi$ from the $\beta$ associated with $n_{it}$, say $\beta_n$, which is an element of $\kappa_{it}$ in the utility function; only $\beta_n + \pi$ is identified.

- Joint normality is sufficient to identify the wage parameters, $\gamma$ and $\sigma_\eta^2$, as well as $(\sigma_\eta^2 - \sigma_{\varepsilon\eta})/\sigma_\xi$ (Heckman (1978)).

- Data on work choices identify $\gamma/\sigma_\xi$ and $\beta/\sigma_\xi$.

- To identify $\sigma_\xi$, note that there are three possible types of variables that appear in the likelihood function:

- variables that appear only in $z$ (in the wage function)
- variables that appear only in $\kappa$ (in the value of leisure)
- variables that appear in both $\kappa$ and $z$.

Having identified the parameters of the wage function (the $\gamma' s$), the identification of $\sigma_\xi$ (and thus also $\sigma_{\varepsilon\eta}$) requires the existence of at least one variable that appears only in the wage equation.

Goal 1: As in the NS approaches, there must be an exclusion restriction, in particular, a variable in $z_{it}$ that is not in $\kappa_{it}$.

Goal 2. It is possible to determine the effect on participation of a change in any of the variables within and outside of the range of the data.

Goal 3: It is possible to identify $\beta_n + \pi$. Suppose that a policy maker is considering implementing a child care subsidy in which the couple is provided a subsidy of $\tau$ dollars for each child if the wife works

The couple's budget constraint with the subsidy is

$$c_{it} = w_{it} d_{it} + y_{it} - (\pi - \tau) d_{it} n_{it}, \tag{17}$$

where $(\pi - \tau)$ is the net (of subsidy) cost of child care. With the subsidy, the probability that the woman works is

$$\Pr(d_{it} = 1 | \Omega_{it}^{-}, \tau) = \Phi\left(\frac{z_{it}\gamma - \kappa_{it}\beta - (\beta_n + \pi - \tau) n_{it}}{\sigma_{\xi}}\right), \tag{18}$$

where $\Phi$ is the standard normal cumulative.

Given identification of $\beta_n + \pi$, to predict the effect of the policy on participation, that is, the difference in the participation probability when $\tau$ is positive and when $\tau$ is zero, it is necessary to have identified $\sigma_\xi$.

Government outlays on the program would be equal the subsidy amount times the number of women with young children who take-up the subsidy (work).

The policy effect is estimated without direct policy variation, i.e., we do not need to observe families with and without the subsidy program.

- What was critical for identification is (exogenous) variation in the wage (independent of preferences).

- The child care cost is a tax on working that is isomorphic to a tax on the wage. Wage variation, independent of preferences, provides *policy-relevant* variation.

# Summary

- Testing the prediction that participation rises with the wage offer requires an exclusion restriction regardless of the approach, because of the non-observability of wage offers for those that choose not to work.

- With respect to goal 3, because of the subsidy acts like a wage tax, the effect of the subsidy can be calculated by comparing participation rates of women with a given wage to women with a wage augmented by $\pi n_{it}$ (see Ichimura and Taber (2002) and Todd and Wolpin).

- Parametric approach allows extrapolation outside of the sample range of the variables whereas nonparametric approaches do not.

-The P-S approach enables the researcher to perform counterfactual exercises, subsidizing the price of child care in the example, even in the absence of variation in that child care price.

# Dynamic Models

- In the static model, there was no connection between the current participation decision and future utility.
- One way, among many, to introduce dynamics is through human capital accumulation on the job. Suppose that the woman's wage increases with actual work experience, $h$, as skills are acquired through learning by doing.

$$w_{it} = z_{it}\gamma_1 + \gamma_2 h_{it} + \eta_{it}, \tag{19}$$

where $h_{it} = \sum_{\tau=1}^{\tau=t-1} d_{i\tau}$ is work experience at the start of period $t$.

Work experience, $h_{it}$, evolves according to

$$h_{it} = h_{i,t-1} + d_{i,t-1} \tag{20}$$

where $h_{i1} = 0$.
-For now, assume the evolution of other elements of the state space is non-stochastic.
-Assume that the preference shock ($\varepsilon_{it}$) and the wife's wage shock ($\eta_{it}$) are distributed joint normal and are mutually serially independent, that is, $(f(\varepsilon_{it}, \eta_{it} | \varepsilon_{it-1}, \eta_{it-1,\ldots,}\varepsilon_{i1}, \eta_{i1}) = f(\varepsilon_{it}, \eta_{it}))$.

- Assume that the couple maximizes expected present discounted value of remaining lifetime utility at each period starting from an initial period, $t = 1$, and ending at period $T$, the assumed terminal decision period.

-For illustrative purposes, assume wife retires at $T + 1$ and that value function at $T + 1$ is normalized to zero.

- Letting $V_t(\Omega_{it})$ be the maximum expected present discounted value of remaining lifetime utility at $t = 1, ..., T$ given the state space and discount factor $\delta$,

$$V_t(\Omega_{it}) = \max_{d_{it}} E\left\{\sum_{\tau=t}^{\tau=T} \delta^{\tau-t}[U^1_{i\tau}d_{i\tau} + U^0_{i\tau}(1-d_{i\tau})]|\Omega_{i\tau}\right\}. \quad (21)$$

- The state space at $t$ consists of the same elements as in the static model augmented to include $h_{it}$.

The value function ($V_t(\Omega_{it})$) can be written as the maximum over the two alternative-specific value functions, $V_t^k(\Omega_{it})$, $k \in \{0,1\}$

$$V_t(\Omega_{it}) = \max(V_t^0(\Omega_{it}), V_t^1(\Omega_{it})) \tag{22}$$

each of which obeys the Bellman equation

$$\begin{aligned} V_t^k(\Omega_{it}) = & \; U_{it}^k(\Omega_{it}) + \delta E[V_{t+1}(\Omega_{i,t+1})|\Omega_{it}, d_{it} = k] \;\; \text{for } t < T \\ = & \; U_{iT}^k(\Omega_{iT}) \;\; \text{for } t = T. \end{aligned} \tag{23}$$

The expectation is taken over the distribution of the random components of the state space at $t+1$, $\varepsilon_{i,t+1}$ and $\eta_{i,t+1}$, conditional on the state space elements at $t$.

The latent variable in the dynamic case is the difference in alternative-specific value functions, $V_t^1(\Omega_{it}) - V_t^0(\Omega_{it})$:

$$\begin{aligned} v_t^*(\Omega_{it}) &= z_{it}\gamma_1 + \gamma_2 h_{it} - \pi n_{it} - \kappa_{it}\beta - \varepsilon_{it} + \eta_{it} \\ &+ \delta\{[E[V_{t+1}(\Omega_{i,t+1})|\Omega_{it}, d_{it} = 1] - [E[V_{t+1}(\Omega_{i,t+1})|\Omega_{it}, d_{it} = 0]\} \\ &= \xi_{it}^*(\Omega_{it}^-) + \xi_{it}. \end{aligned}$$

To calculate these alternative-specific value functions, note first that $\Omega^-_{i,t+1}$, the observable part of the state space at $t+1$, is fully determined by $\Omega^-_{it}$ and the choice at $t, d_{it}$. Thus, one needs to be able to calculate $E[V_{t+1}(\Omega_{i,t+1})|\Omega_{it}, d_{it})$ at all values of $\Omega^-_{i,t+1}$ that may be reached from the state space elements at $t$ and a choice at $t$. A full solution of the dynamic programming problem consists then of finding $EV_\tau(\Omega_{i\tau}) = E\max[(V^0_\tau(\Omega_{i\tau}), V^1_\tau(\Omega_{i\tau}))]$ for all values of $\Omega^-_{i\tau}$ at all $\tau = 2, \ldots T$. We denote this function by $E\max(\Omega^-_{it})$ or $E\max_t$ for short.

- In this finite horizon model, the solution method is by backwards recursion.
- First, we need to assume something about how the exogenous observable state variables evolve, that is, $z_{it}, n_{it}, \kappa_{it}$. For simplicity for now, assume $z_{it} = z_i$ and $\kappa_{it} = \kappa_i$.
- Number of young children obviously not constant over the life cycle, but assume that the woman is old enough in the decision period so that the evolution of $n_{it}$ is non-stochastic. In general, if $n_{it}$ is the number of children under 6, then the ages of the young children enter the state space.

- To calculate the alternative-specific value functions at period $T-1$ for each element of $\Omega^-_{i,T-1}$, we need to calculate $E\max_T$.
- Using the fact that, under normality,

$E(\varepsilon_{iT}|\xi_{iT} < -\xi^*_{iT}(\Omega^-_{iT})) = -\frac{\sigma_{\varepsilon\xi}}{\sigma_\xi}\frac{\phi(-\xi^*_{iT}(\Omega^-_{iT}))}{\Phi(-\xi^*_{iT}(\Omega^-_{iT}))}$ and

$E(\eta_{iT}|\xi_{iT} \geq -\xi^*_{iT}(\Omega^-_{iT}) = \frac{\sigma_{\eta\xi}}{\sigma_\xi}\frac{\phi(-\xi^*_{iT}(\Omega^-_{iT}))}{1-\Phi(-\xi^*_{iT}(\Omega^-_{iT}))}$, we get

$$\text{Emax}_T = y_{iT} + (\kappa_i\beta)\Phi(-\xi^*_{iT}(\Omega^-_{iT})) + (z_i\gamma_1 + \gamma_2 h_{iT} - \pi n_{iT}) \times (1 - \Phi(-\xi^*_{iT}(\Omega^-_{iT})) + \sigma_\xi \phi(-\xi^*_{iT}(\Omega^-_{iT})).$$

-Uses the fact that for any two random variables $u$ and $v$,

$E\max(u,v) = E(u|u>v)\Pr(u>v) + E(v|v>u)\Pr(v>u)$.

-This expression requires an integration (the normal cdf) which has no closed form and must be computed numerically.

- The RHS of is a function of $y_{iT}, z_i, \kappa_i, n_{iT}$ and $h_{iT}$.
- Given a set of model parameters, the $E\max_T$ function takes on a scalar value for each element of its arguments.
- Noting that $h_{iT} = h_{i,T-1} + d_{i,T-1}$, and being explicit about the elements of $E\max_T$, the alternative-specific value functions at T-1 are (dropping the $i$ subscript for convenience):

$$\begin{aligned} V^0_{T-1}(\Omega_{T-1}) &= \; y_{T-1} + \kappa\beta + \varepsilon_{T-1} + \delta E\max(y_T, z, \kappa, n_T, h_{T-1}), \\ V^1_{T-1}(\Omega_{T-1}) &= y_{T-1} + z\gamma_1 + \gamma_2 h_{T-1} - \pi n_{T-1} + \eta_{T-1} \\ &\quad + \delta E\max(y_T, z, \kappa, n_T, h_{T-1} + 1). \end{aligned}$$

Thus,

$$\begin{aligned} v^*_{T-1}(\Omega_{i,T-1}) &= z\gamma_1 + \gamma_2 h_{T-1} - \pi n_{T-1} - \kappa\beta - \varepsilon_{T-1} + \eta_{T-1} \\ &+ \delta\left\{E\max(y_T, z, \kappa, n_T, h_{T-1}+1) - E\max(y_T, z, \kappa, n_T, h_{T-1})\right\} \\ &= \xi^*_{T-1}(\Omega^-_{T-1}) + \xi_{T-1}. \end{aligned}$$

Because $y_T$ enters both $E\max(y_T, z, \kappa, n_T, h_{T-1}+1)$ and $E\max(y_T, z, \kappa, n_T, h_{T-1})$ additively, it drops out of $\xi^*_{T-1}(\Omega^-_{T-1})$ and thus out of $v^*_{T-1}$

- To calculate the T-2 alternative-specific value functions, will need to calculate $E\max_{T-1}$.

$$
\begin{aligned}
E\max_{T-1} = y_{T-1} &+ (\kappa\beta + \delta E\max(y_{T-1}, z, \kappa, n_T, h_{T-1}))\Phi(-\xi^*_{T-1}(\Omega^-_{T-1})) \\
&+ (z\gamma_1 + \gamma_2 h_{T-1} - \pi n_{T-1} + \delta E\max(y_{T-1}, z, \kappa, n_T, h_{T-1}+1))(1 - \Phi(-\xi^*_{T-1}(\Omega^-_{T-1})) \\
&+ \sigma_\xi \phi(-\xi^*_{T-1}(\Omega^-_{T-1})).
\end{aligned}
$$

The RHS is a function of $y_{T-1}, z, \kappa, n_{T-1}, n_T$ and $h_{T-1}$.

- As with $E\max_T$, given a set of model parameters, the $E\max_{T-1}$ function takes on a scalar value for each element of its arguments.
- The alternative-specific value functions at T-1 and the latent variable function are:

$$
\begin{aligned}
V^0_{T-2}(\Omega_{T-2}) &= \ y_{T-2} + \kappa\beta + \varepsilon_{T-2} + \delta E\max(y_{T-1}, z, \kappa, n_{T-1}, n_T, h_{T-2}), \qquad (24)\\
V^1_{T-2}(\Omega_{T-2}) &= y_{T-2} + z\gamma_1 + \gamma_2 h_{T-2} - \pi n_{T-2} + \eta_{T-2}\\
&\quad + \delta E\max(y_{T-1}, z, \kappa, n_{T-1}, n_T, h_{T-2}+1),\\
v^*_{T-2}(\Omega_{T-2}) &= z\gamma_1 + \gamma_2 h_{.T-2} - \pi n_{T-2} - \kappa\beta - \varepsilon_{T-2} + \ \eta_{i,T-2}\\
&\quad + \delta\left\{E\max(y_{T-1}, z, \kappa, n_{T-1}, n_T, h_{T-2}+1) - E\max(y_{T-1}, z, \kappa, n_{T-1}, n_T, h_{T-2})\right\}\\
&= \xi^*_{T-2}(\Omega^-_{T-2}) + \xi_{T-2}.
\end{aligned}
$$

As at $T$, $y_{T-1}$ drops out of $\xi^*_{T-2}(\Omega^-_{T-2})$ and thus $v^*_{T-2}$.

- Continue to solve backwards in this fashion.
- The full solution of the dynamic programming problem is the set of $E\max_t$ functions for all $t$ from $t = 1, .., T$. - These $E\max_t$ functions provide all of the information necessary to calculate the cut-off values, the $\xi_t^*(\Omega_t^-)'s$ that are the inputs into the likelihood function.

- Estimation of the dynamic model requires that the researcher have data on work experience, $h_{it}$.
- Assume that the researcher has longitudinal data for $I$ married couples and denote by $t_{1i}$ and $t_{Li}$ the first and last periods of data observed for married couple $i$.
- Note that $t_{1i}$ need not be the first period of marriage (although it may be, subject to the marriage occurring after the woman's fecund period) and $t_{Li}$ need not be the last (although it may be).

- Denoting $\theta$ as the vector of model parameters, the likelihood function is:

$$L(\theta; data) = \Pi_{i=1}^{i=I} \Pi_{\tau=t_{1i}}^{\tau=t_{Li}} \Pr(d_{i\tau} = 1, w_{i\tau} | \Omega_{i\tau}^{-})^{d_{i\tau}} \Pr(d_{i\tau} = 0 | \Omega_{i\tau}^{-})^{1-d_{i\tau}}, \quad (25)$$

where $\Pr(d_{i\tau} = 1, w_{i\tau} | \Omega_{i\tau}^{-}) = \Pr( \xi_{i\tau} \geq -\xi_{i\tau}^{*}(\Omega_{i\tau}^{-}),$
$\eta_{i\tau} = w_{i\tau} - z_{i\tau}\gamma_1 - \gamma_2 \; h_{i\tau})$ and
$\Pr(d_{i\tau} = 0 | \Omega_{i\tau}^{-}) = 1 - \Pr( \xi_{i\tau} \geq -\xi_{i\tau}^{*}(\Omega_{i\tau}^{-}).$

- If the structure does not yield an additive (composite) error, the latent variable function becomes $v_t^*(\Omega_{it}^-, \eta_{it}, \varepsilon_{it})$.

- Calculating the joint regions of $\eta_{it}$, $\varepsilon_{it}$ that determine the probabilities that enter the likelihood function and that are used to calculate the $E\max(\Omega_{it}^-)$ function must, in that case, be done numerically.

# Likelihood Estimation

Given joint normality of $\varepsilon$ and $\eta$, the likelihood function is analytic

$$L(\theta; data) = \prod_{i=1}^{i=I} \prod_{\tau=t_{1i}}^{\tau=t_{Li}} \left\{ \left[ 1 - \Phi\left( \frac{-\xi_\tau^*(\Omega_{i\tau}^-) - \rho \frac{\sigma_\xi}{\sigma_\eta} \eta_{i\tau}}{\sigma_\xi (1-\rho^2)^{\frac{1}{2}}} \right) \right] \frac{1}{\sigma_\eta} \phi\left( \frac{\eta_{i\tau}}{\sigma_\eta} \right) \right\}^{d_{i\tau}} \tag{26}$$

$$\times \left\{ \Phi\left( \frac{-\xi_\tau^*(\Omega_{i\tau}^-)}{\sigma_\xi} \right) \right\}^{1-d_{i\tau}} .$$

where $\eta_{i\tau} = w_{i\tau} - z_{i\tau}\gamma_1 - \gamma_2 h_{i\tau}$ and where $\rho$ is the correlation coefficient between $\xi$ and $\eta$.

# Likelihood Estimation

- Estimation proceeds by iterating between the solution of the dynamic programming problem and the likelihood function for alternative sets of parameters. - Maximum likelihood estimates are consistent, asymptotically normal and efficient.

- Given the solution of the dynamic programming problem for the cut-off values, the $\xi^*_{it}(\Omega^-_{it})$'s, the estimation of the dynamic model is in principle no different than the estimation of the static model.

- There is the additional discount factor to be estimation, plus the additional assumptions on how households forecast future unobservables.

- The practical difference is the computational effort of having to solve the dynamic programming problem in each iteration on the model parameters in maximizing the likelihood function.

- Identification of model parameters requires the same exclusion restriction as in the static case, at least one variable in the wage equation that does not affect the value of leisure.

- Work experience, $h_{it}$, would serve that role if it does not also enter into the value of leisure ($\kappa$).

# Identification of the discount factor

- the difference in the future component of the expected value functions under the two alternatives is in general a non-linear function of the state variables and depends on the same set of parameters as in the static case.

Rewriting the latent variable equation as:

$$v_t^*(\Omega_{it}) = z_i\gamma_1 + \gamma_2 h_{it} - \pi n_{it} - \kappa_i\beta + \delta W_{t+1}(\Omega_{it}^-) - \varepsilon_{it} + \eta_{it}, \quad (27)$$

where $W(\cdot)$ is the difference in the future component of the expected value functions, the non-linearities in $W_{t+1}$ that arise from the distributional and functional form assumptions may be sufficient to identify the discount factor.

- identification of the model parameters implies that all three research goals (described previously) can be met.
- A quantitative assessment of the counterfactual child care subsidy is feasible.
- the effect of such a subsidy will differ from a static model as any effect of the subsidy on the current participation decision will be transmitted to future participation decisions through the change in work experience and future wages.

- If a surprise (permanent) subsidy were introduced at time $t$, the effect of the subsidy on participation at $t$ would require that the couple's dynamic programming problem be resolved with the subsidy from $t$ to $T$ and the solution compared to that without the subsidy.
- A pre-announced subsidy to take effect at $t$ would require that the solution be obtained back to the period of the announcement because, given the dynamics, such a program would have effects on participation starting from the date of the announcement.

# Independent additive type-1 extreme value errors:

- When shocks are additive and come from independent type-1 extreme value distributions, as first noted by Rust (1987), the solution to the dynamic programming problem and the choice probability both have closed forms and do not require a numerical integration as in the additive normal error case.
- The cdf of an extreme value random variable $u$ is $\exp(-e^{-\frac{u}{\rho}})$ and has mean equal to $\rho\gamma$, where $\gamma$ is Euler's constant, and variance $\frac{\pi^2\rho^2}{6}$.

Under the extreme value assumption, it can be shown that for period $t = T$ (dropping the $i$ subscript for convenience)

$$\Pr(d_T = 1|\Omega_T^-) = \exp\left(\frac{z\gamma_1 + \gamma_2 h_T - \pi n_T - \kappa\beta}{\rho}\right)\left(1 + \exp(\frac{z\gamma_1 + \gamma_2 h_T - \pi n_T - \kappa\beta}{\rho})\right)^{-1}$$

$$E\max{}_T = \rho\left\{\gamma + \log\left[\exp\left(\frac{y_T + z\gamma_1 + \gamma_2 h_T - \pi n_T}{\rho}\right) + \exp\left(\frac{y_T + \kappa\beta}{\rho}\right)\right]\right\}$$

$$= \rho\left\{\gamma + \frac{y_T + z\gamma_1 + \gamma_2 h_T - \pi n_T}{\rho} - \log(\Pr(d_T = 1|\Omega_T^-)\right\}$$

# f

or $t < T$,

$$\Pr(d_t = 1|\Omega_t^-) = \frac{\exp(\frac{z\gamma_1+\gamma_2 h_t-\pi n_t-\kappa\beta+\delta\{E\max_t(y_{t+1},z,\kappa,\widetilde{n}_{t+1},h_t+1)-E\max_t(y_{t+1},z,\kappa,\widetilde{n}_{t+1},h_t)\}}{\rho})}{1+\exp(\frac{z\gamma_1+\gamma_2 h_T-\pi n_t-\kappa\beta+\delta\{E\max_t(y_{t+1},z,\kappa,\widetilde{n}_{t+1},h_t+1)-E\max_t(y_{t+1},z,\kappa,\widetilde{n}_{t+1},h_t)\}}{\rho}} \tag{28}$$

$$E\max{}_t = \rho\left\{\gamma + \log\left[\exp\left(\frac{V_t^1(\Omega_t^-)}{\rho}\right) + \exp\left(\frac{V_t^0(\Omega_t^-)}{\rho}\right)\right]\right\} \tag{29}$$

$$= \rho\left\{\gamma + \frac{y_t + z\gamma_1 + \gamma_2 h_t - \pi n_t + \delta E\max(y_{t+1}, z, \kappa, \widetilde{n}_{t+1}, h_t + 1)}{\rho} - \log(\Pr(d_t = 1|\Omega_{it}^-)\right\}$$

where we have let $\widetilde{n}_{t+1}$ stand for the vector of $n_{t+1}, \ldots, n_T$ values.

- The solution, as in the case of normal errors, consists of calculating the $E\max_t$ functions by backwards recursion.
- Unlike for normal errors, the $E\max_t$ functions and the choice probabilities have closed form solutions.

- The extreme value assumption is problematic in the labor force participation model. For there to be a closed form solution to the DCDP problem, both the preference shock and the wage shock must be extreme value.
- Could modify the problem so that (a) the wife's wage offer is not observed at the time that the participation decision is made or (b) the wage is deterministic (but varies over time and across women due to measurement error).
-Then, by adding an independent type-1 extreme value error to the utility when the wife works, the participation decision rule will depend on the difference in two extreme value taste errors, which leads to the closed form expressions given above.

- In either case, there is no longer a selection issue with respect to observed wages.
- Because the observed wage shock is independent of the participation decision, the wage parameters can be estimated by adding the wage density to the likelihood function for participation and any distributional assumption, (e.g. log normality) can be assumed.
- Whether the model assumptions necessary to take advantage of the computational gains from adopting the extreme value distribution are warranted raises the issue how models should be judged and which model is "best."

# Unobserved State Variables:

- Would like to relax the assumption of shocks being independent of past shocks.
- Common to assume that shocks have a permanent-transitory structure (HeckmanSinger, 1981, KeaneWolpin, 1997), where the permanent component takes on a discrete number of values and follows a joint multinomial distribution:

$$\varepsilon_{it} = \sum_{m^h=1}^{M} \sum_{m^w=1}^{M} \lambda_{1m^h m^w} 1(type^h = m^h, type^w = m^w) + \omega_{1it},$$

$$\eta_{it} = \sum_{m^w=1}^{M} \lambda_{2m} 1(type^w = m^w) + \omega_{2it}.$$

where there are $M$ types each of husbands ($h$) and wives ($w$), and thus $M^2$ couple types and where $\omega_{1it}$ and $\omega_{2it}$ are joint normal and iid over time.

- A couple is assumed to know their own and their spouse's type, so the state space is augmented by the husband's and wife's type (or could impose assumption that they are the same types).

- Even though researcher does not know types, it is convenient to add them to the state variables in what we previously defined as the observable elements of the state space, $\Omega_{it}^{-}$.

- The dynamic programming problem must be solved for each couple's type.

# Likelihood with unobserved types

- The likelihood function modified to account for the unobserved types.

- Letting $L_{(m^w,m^h)}$ be the likelihood function for a type $(m^w, m^h)$ couple.

- The sample likelihood is the product over individuals of the type probability weighted sum of the type-specific likelihoods:

$$\prod_i L^i = \sum_{m^w=1}^{M} \sum_{m^h=1}^{M} \pi_{m^w m^h} L^i_{(m^w,m^h)}. \tag{30}$$

-Another possibility is to assume the joint error process follows an ARIMA process. Suppose the errors follow a AR(1)process:

$\varepsilon_{i,t} = \rho_e \varepsilon_{i,t-1} + \omega_{1it}$ and $\eta_{it} = \rho_\eta \eta_{i,t-1} + \omega_{2it}$,
where $\omega_{1it}$ and $\omega_{2it}$ are joint normal and iid over time.

- Consider again the alternative-specific value functions at $t$, where we now explicitly account for the evolution of the shocks:

$$\begin{aligned} V_t^k(\Omega_{it}^-, \varepsilon_{it}, \eta_{it}) &= U_{it}^k(\Omega_{it}) + \delta E[V_{t+1}(\Omega_{i,t+1}^-, \varepsilon_{it+1}, \eta_{it+1}) | \Omega_{it}^-, \varepsilon_{it}, \eta_{it}, d_{it} = k] \qquad (31) \\ &= U_{it}^k(\Omega_{it}) + \delta E[V_{t+1}(\Omega_{i,t+1}^-, \rho_e \varepsilon_{it} + \omega_{1it+1}, \rho_\eta \eta_{it} + \omega_{2it+1}) | \Omega_{it}^-, \varepsilon_{it}, \eta_{it}, d_{it} = k] \end{aligned}$$

the integration is taken over the joint distribution of $\omega_{1it+1}$ and $\omega_{2it+1}$

- The $E\max_{t+1}$ function includes not only $\Omega^{-}_{i,t+1}$, as previously specified, but also the shocks at $t$, $\varepsilon_{it}$ and $\eta_{it}$.

- Serial correlation augments the state space that enters the $E\max_t$ functions.

- The main complication is that these state space elements, unlike those we had so far, are continuous variables.

# The Curse of Dimensionality:

- The solution of the dynamic programming problem required that the $E\max_t$ functions be calculated for each point in the state space.

- If $z$ and $\kappa$ take on only a finite number of discrete values (e.g., years of schooling, number of children), the solution simply involves solving for the $E\max_t$ functions at each point in the state space

- If either $z$ or $\kappa$ contains a continuous variable (or if the shocks follow an ARIMA process), one cannot solve the dynamic programming problem at every state point.

- One could also imagine making the model more complex in ways that would increase the number of state variables and hence the size of the state space (e.g. including ages of children).

- The state space grows exponentially with the number of state variables. This is the *curse of dimensionality* first associated with Bellman (1957).

- Estimation requires solving the dynamic programming problem many times for each trial parameter vector considered in the search for the maximum of the likelihood function (and perhaps at many nearby parameter vectors, in order to obtain gradients).

- Two main ways to make calculation of the DP problem feasible: (a) keep the model simple so state space is small, or (b) abandon "exact" solutions to DP problems in favor of approximate solutions that can be obtained with greatly reduced computational time.

There are three main approximate solution methods that have been discussed in the literature:

1. *Discretization:* discretize the continuous variables and solve for the $E\max_t$ functions on the grid of discretized values. Either
(i) modify the law of motion for the state variables so they stay on the discrete grid, or
(ii) use a method to interpolate between grid points.

2. *Approximation and interpolation* of the $E\max_t$ functions: approach was originally proposed by Bellman, Kalaba and Kotkin (1963) and extended to the type of models generally of interest by Keane and Wolpin (1994).

-Applicable when the state space is large either due the presence of continuous state variables or a large number of discrete state variables (or both).

- The $E\max_t$ functions are evaluated at a subset of the state points and some method of interpolation is used to evaluate $E\max_t$ at other values of the state space.

- Requires that the $E\max_t$ interpolating functions be parametrically specified.

3. *Randomization*: developed by Rust (1997). Applicable when the state space is large due the presence of continuous state variables.

- Requires that choice variables be discrete and state variables be continuous.
- Rust (1997) shows that solving a random Bellman equation can break the curse of dimensionality in the case of DCDP models in which the state space is continuous and evolves stochastically, conditional on the alternative chosen.
- Suppose that we modeled work experience as a continuous random variable with conditional density function $p(h_{t+1}|h_t, d_t) = p(h_t + jI(d_t = 1) - jI(d_t = 0)|h_t, d_t)$ where $j$ is random variable indicating the extent to which working probabilistically augments work experience or not working depletes effective work experience (due to depreciation of skills).

The random Bellman equation (ignoring $z$ and $\kappa$), is

$$\widehat{V}_{Mt}(h_t) = \max_{d_t} \left[ U_t^{d_t}(h_t) + \frac{\delta}{M} \sum_{m=1}^{M} \widehat{V}_{M,t+1}(h_{t+1,m}|h_t, d_t) p(h_{t+1,m}|h_t, d_t) \right], \tag{32}$$

where $[h_{t+1,1}, \ldots, h_{t+1,M}] = [h_1, \ldots, h_M]$ are $M$ randomly drawn state space elements.

- The approximate value function $\widehat{V}_{Mt}(h_t)$ converges to $V_t(h_t)$ as $M \to \infty$ at a $\sqrt{M}$ rate.

- Notice that this also true if $(h_t)$ is a vector of state variables. -

The above approach only delivers a solution for the value functions on the grid $[h_1, ..., h_M]$, but evaluating the likelihood will usually require values at other points.

- A key point is that $\widehat{V}_{Mt}(h_t)$ is (in Rust's terminology) self-approximating.

- Suppose we wish to construct the alternative specific value function $\widehat{V}_{M_t}^{d_t}(h_t)$ at a point $h_t$ that is not part of the grid $[h_1, ..., h_M]$. Form:

$$\widehat{V}_{M_t}^{d_t}(h_t) = U_t^{d_t}(d_t) + \delta \sum_{m=1}^{M} \widehat{V}_{M_t}(h_m) \frac{p(h_m | h_t, d_t)}{\sum_{k=1}^{M} p(h_k | h_t, d_t)}.$$

- Because any state space element at $t+1$ can be reached from any element at $t$ with some probability given by $p(\cdot | \ h_t, d_t)$, the value function at $t$ can be calculate at any element of the state space at $t$.

# Comparison of different approaches

- For any given application with a given number of state variables it is an empirical question whether methods based on discretization, approximation/interpolation or randomization will produce a more accurate approximation in given computation time.

- Stinebrickner(2000) compares several approximation methods in the context of a DCDP model with serially correlated shocks.

- more work is needed to understand which methods perform best and in what contexts.

# The Multinomial Dynamic Discrete Choice Problem

Extend that model to allow for:

(i) additional choices

(ii) non-additive errors

(iii) general functional forms and distributional assumptions.

- One can simply allow the number of mutually exclusive alternatives, and thus the number of alternative-specific value functions to be greater than two.

-For example, if there are $K > 2$ mutually exclusive alternatives, there will be $K - 1$ latent variable functions (relative to one of the alternatives, arbitrarily chosen).

- Consider the extension of DCDP models to the case with multiple discrete alternatives by augmenting the dynamic labor force participation model to include a fertility decision in each period.

- In addition, allow the couple to choose among four labor force alternatives for the wife.

- Drop the assumption that errors are additive and normal (which would be unusual for wages)

# Richer Model of LFP and Fertility Choices

- Consider a sample of white married women (in their first marriage) taken from the 1979-2004 rounds of the NLSY79. Ages at marriage range from 18 to 43, with 3/4ths of these first marriages occurring before the age of 27.
- Adopt, as is common in labor supply models, a discrete period length of a year.
- The participation measure consists of four mutually exclusive and exhaustive alternatives, working less than 500 hours during a calendar year ($d_{it}^0 = 1$), working between 500 and 1499 hours ($d_{it}^1 = 1$), working between 1500 and 2499 hours ($d_{it}^2 = 1$) and working more than 2500 hours ($d_{it}^3 = 1$).
- Fertility is the dichotomous variable indicating whether or not the woman had a birth during the calendar year.

# Approximate decision rules

- It is often useful to estimate the relationship between the choice variables (hours worked, fertility) and the state variables, just to make sure that they are related in the way you might expect.

-Let the married couple's per-period utility flow include consumption ($c_{it}$), a per-period disutility from each working alternative and a per-period utility flow from the stock of children ($N_{it}$).

-The stock of children includes a newborn, that is a child born at the beginning of period $t$ ($n_{it} = 1$).

$$U_{it} = U(c_{it}, d_{it}^1, d_{it}^2, d_{it}^3, N_{it}; \varepsilon_{it}^1 d_{it}^1, \varepsilon_{it}^2 d_{it}^2, \varepsilon_{it}^3 d_{it}^3, \varepsilon_{it}^N N_{it}) \qquad (33)$$

where the $\varepsilon_{it}^1$, $\varepsilon_{it}^2, \varepsilon_{it}^3$, and $\varepsilon_{it}^N$are time-varying preference shocks that are assumed to be mutually serially uncorrelated.

Because the stock of children at $t$ includes a newborn, the utility of a newborn is subject to the $\varepsilon_{it}^{N}$ shock.

Allowing for unobserved heterogeneity, the type specification is

$$\varepsilon_{it}^{j} = \sum_{m^h=1}^{M} \sum_{m^w=1}^{M} \lambda_{1m}^{j} 1(type^h = m^h, type^w = m^w) + \omega_{1it}^{j}, j = 1, 2, 3, N, \tag{34}$$

where the $\omega^{j\prime}s$ are mutually serially independent shocks.

- The household budget constraint incorporates a cost of avoiding a birth (contraceptive costs, $b_0$, which, for biological reasons, will be a function of the wife's age (her age at marriage, $a_0^w$, plus the duration of marriage, $t$) and (child) age-specific monetary costs of supplying children with consumption goods ($b_{1k}$) and with child care if the woman works ($b_{2k}$ per work hour).

- Household income=sum of husband's earnings ($y_{it}$)and wife's earnings, the product of an hourly wage ($w_{it}$) and hours worked (1000 hours if $d_{it}^1 = 1$, 2000 hours if $d_{it}^2$, 3000 hours $d_{it}^3 = 1$).

$$
\begin{aligned}
c_{it} = y_{it} + w_{it}(1000d_{it}^1 + 2000d_{it}^2 + 3000d_{it}^3) \\
- b_0(a_0^w + t)(1 - n_{it}) - \sum_{k=1}^{K} b_{1k} N_{kit} \\
- \sum_{k=1}^{K} b_{2k} N_{kit}(1000d_{it}^1 + 2000d_{it}^2 + 3000d_{it}^3)
\end{aligned}
$$

where $N_{kit}$ are the number of children in $K$ different age classes, e.g., 0-1, 2-5, etc.

The wife's wage offer function depends on her level of human capital, $\Psi_{it}$, assumed to be a function of the wife's completed schooling $(S_i^w)$,, the number of hours worked up to $t$, $E_{it}$ and on the number of hours worked in the previous period:

$$\log w_{it} = \sum_{j=1}^{3} \log r^j d_{it}^j + \log \Psi_{it}(S_i^w, E_{it}, d_{it-1}^1, d_{it-1}^2, d_{it-1}^3; \eta_{it}^w),$$

$$\eta_{it}^w = \sum_{m^w=1}^{M} \lambda_{2m^w} 1(type^w = m^w) + \omega_{2it}^w$$

where the $r^j$ are (assumed to be time-invariant) competitively determined skill rental prices that may differ by hours worked and $\eta_{it}^w$ is a time varying shock to the wife's human capital following a permanent (discrete type)-transitory scheme.

Husband's earnings (assumed to work full-time):

$$\log y_{it} = \log r^h + \log \Psi_{it}^h(S_i^h, a_t^h; \eta_{it}^h),$$

$$\eta_{it}^h = \sum_{m^h=1}^{M} \lambda_{2m^h}^h 1(type^h = m^h) + \omega_{2it}^y.$$

where $S_i^h$ is the husband's schooling and $a_t^h = a_0^h + t$ is his age at $t$ (his age at marriage plus $t$).

The time-varying *state variables*, the stock of children (older than one) of different ages, the total stock of children and work experience, evolve according to:

$$N_{2it} = \sum_{j=t-1}^{t-5} n_{ij}; N_{3it} = \sum_{j=t-6}^{t-17} n_{ij}; N_{it} = N_{it-1} + n_{it}, \tag{35}$$

$$E_{it} = E_{it-1} + +1000 d^1_{it-1} + 2000 d^2_{it-1} + 3000 d^3_{it-1}. \tag{36}$$

The state variables in $\Omega_t^-$, augmented to include type, consist of the stock of children (older than one) of different ages, the wife's work experience and previous period work status, the husband's and wife's age at marriage, the husband and wife's schooling levels and the couple's type.

The *choice set* during periods when the wife is fecund, assumed to have a known terminal period ($t_m$), consists of the four work alternatives plus the decision of whether or not to have a child.

-There are thus eight mutually exclusive choices, $d_{it}^{hn} = \{d_{it}^{00}, d_{it}^{10}, d_{it}^{20}, d_{it}^{30}, d_{it}^{01}, d_{it}^{11}, d_{it}^{21}, d_{it}^{31} : t = 1, ..., t_m - 1\}$, where the first superscript refers to the work choice ($h = \{0, 1, 2, 3\}$) and the second to the fertility choice ($n = \{0, 1\}$)

-When the wife is no longer fecund, $n_{it} = 0$ and the choice set consists only of the four mutually exclusive alternatives, $d_{it}^{hn} = \{d_{it}^{00}, d_{it}^{10}, d_{it}^{20}, d_{it}^{30} : t = t_m, ..., T\}$.

- Couples choose the alternative at each $t$ that maximizes the remaining expected discounted value of lifetime utility.

- Defining $U_{it}^{hn}$ to be the contemporaneous utility flow for the work and fertility choice, the value functions are:

$$V_t^{hn}(\Omega_{it}) = U_{it}^{hn}(\Omega_{it}) + \delta E[V_{t+1}(\Omega_{i,t+1})|\Omega_{it}^{-}, d_{it}^{hn}] \quad \text{for } t < T, \tag{37}$$

$$= U_{iT}^{hn}(\Omega_{iT}) \quad \text{for } t = T \tag{38}$$

where, letting $\widetilde{V_t^{hn}}$ be the vector of alternative specific value functions relevant at period $t$,

$$V_t(\Omega_{it}) = \max(\widetilde{V_t^{hn}}(\Omega_{it})), \tag{39}$$

and where the expectation is taken over the joint distribution of the preference and income shocks, $f(\omega_{1t}^1, \omega_{1t}^2, \omega_{1t}^3, \omega_{1t}^N, \omega_{2t}^w, \omega_{2t}^y)$.

# Model solution

- The model is solved by backwards recursion.
- The solution requires, as in the binary case, that the $E\max_t$ function be calculated at each state point and for all $t$.
- Here, the $E\max_t$ function is a six-variate integral (over the preference shocks, the wife's wage shock and the husband's earnings shock.
- The state space at $t$ consists of all feasible values of $E_{it}, d^1_{it-1}, d^2_{it-1}, d^3_{it-1}, S^w, S^h, N_{it-1}, N_{kit}$ $(k = 2, 3), a^h_0, a^w_0, type^h, type^w$.

- All of the state variables are discrete and the dimension of the state space is therefore finite, but large.
- To track the number of children in each of the three age groups, it is necessary to keep track of the complete sequence of births. If a woman has 30 fecund periods, the number of possible birth sequences is $2^{30} = 1,073,700,000$.
- Full solution of the dynamic programming problem is infeasible, leaving aside the iterative process necessary for estimation.

- It is thus necessary to use an approximation method for solving for the $E\max_t$ functions.
- Here is an interpolation method based on regression. Consider first the calculation of the $E\max_T$ for any given state space element. At $T$ the woman is no longer fecund, so we need to calculate

$$E\max_T = E_{T-1}\max(U_T^{00}(\widetilde{\omega}), U_T^{10}(\widetilde{\omega}), U_T^{20}(\widetilde{\omega}), U_T^{30}(\widetilde{\omega})), \quad (40)$$

where $\widetilde{\omega}$ is the six-tuple vector of shocks.

- This is a six-variate integration and $E\max_T$ must be calculated numerically. A straightforward method is Monte Carlo integration.
- Letting $\widetilde{\omega}_d$ be the $d^{th}$ random draw, $d = 1, ..., D$, from the joint distribution, $f(\omega_1^1, \omega_1^2, \omega_1^3, \omega_1^N, \omega_2^w, \omega_2^h)$, an estimate of $E\max_T$ at say the $k$th value of the state space in $\Omega_T^-$, $\Omega_{Tk}^-$, is

$$\widehat{E\max}_{Tk} = \frac{1}{D}\sum_{d=1}^{D} \max[U_T^{00}(\widetilde{\omega_d}; \Omega_{Tk}^-), U_T^{10}(\widetilde{\omega_d}; \Omega_{Tk}^-), U_T^{20}(\widetilde{\omega_d}; \Omega_{Tk}^-), U_T^{30}(\widetilde{\omega_d}; \Omega_{Tk}^-)] \tag{41}$$

- Suppose one randomly draws $K_T$ state points (without replacement) and calculates the $\widehat{E\max}_T$ function for those $K_T$ state space elements.
- Can treat these $K_T$ values as a vector of dependent variables in an interpolating regression

$$\widehat{E\max}_{Tk} = g_T(\Omega^{-}_{Tk}; \gamma_T) + \varsigma_T,$$

  where $\gamma_T$ is a time $T$ vector of regression coefficients and $g_T(\cdot;\cdot)$ is a flexible function of state variables.
- With this interpolating function, estimates of the $E\max_T$ function can be obtained at any state point in the set $\Omega^{-}_T$.

- Given $\widehat{E\max}_T$, we can calculate $V^{hn}_{T-1}$ at a subset of the state points in $\Omega^-_{T-1}$.
- Using the $D$ draws from $f(\widetilde{\omega})$, the estimate of $E\max_{T-1}$ at the $k$th state space element is

$$\widehat{E\max}_{T-1,k} = \frac{1}{D}\sum_{d=1}^{D} \max[V^{00}_{T-1}(\widetilde{\omega_d};\Omega^-_{T-1,k}), V^{10}_{T-1}(\widetilde{\omega_d};\Omega^-_{T-1,k}),$$

$$V^{20}_{T-1}(\widetilde{\omega_d};\Omega^-_{T-1,k}), V^{30}_{T-1}(\widetilde{\omega_d};\Omega^-_{T-1,k})] \quad (42)$$

Using the $\widehat{E\max}_{T-1,k}$ calculated for $K_{T-1}$ randomly drawn state points from $\Omega^{-}_{T-1}$ as the dependent variables in the interpolating function:

$$\widehat{E\max}_{T-1,k} = g_{T-1}(\Omega^{-}_{T-1,k};\gamma_{T-1}) + \varsigma_{T-1}, \tag{43}$$

provides estimated values for the $E\max_{T-1}$ function at any state point in the set $\Omega^{-}_{T-1}$.

- Continuing this procedure, we can obtain the interplolating functions for all of the $\widehat{E\max_t}$ functions.
- At some time period, the choice set will include the birth of a child.

# Likelihood estimation

- In the binary case with additive normal errors, the cut-off values for the participation decision, which were the ingredients for the likelihood function calculation, were analytical.
- In the multinomial choice setting, the set of values of the $\boldsymbol{\omega}$ vector that determine optimal choices and that serve as limits of integration in the probabilities associated with the work alternatives that comprise the likelihood function have no analytical form and the likelihood function requires a multivariate integration.
- Maximum likelihood estimation of the model uses simulation methods.

To describe the procedure, let the set of values of $\widetilde{\omega_t}$ for which the $hn^{th}$ choice is optimal at $t$ be denoted by

$S_t^{hn}(\Omega_{it}^{-}) = \{\omega_{1t}^1, \omega_{1t}^2, \omega_{1t}^3, \omega_{1t}^N, \omega_{2t}^w, \omega_{2t}^y | V_t^{hn} = \max(\widetilde{V_t^{hn}})\}$.

Consider the probability that a couple chooses neither to work nor have a child, $h_{it} = 0, n_{it} = 0$, in a fecund period $t < t_m$ :

$$\Pr(h_{it} = 0, n_{it} = 0 | \Omega_{it}^{-}) = \int_{S_t^{00}(\Omega_{it}^{-})} f(\omega_{1t}^1, \omega_{1t}^2, \omega_{1t}^3, \omega_{1t}^N, \omega_{2t}^w, \omega_{2t}^y) d\omega_{1t}^1 d\omega_{1t}^2 d\omega_{1t}^3 d\omega_{1t}^N d\omega_{2t}^w d\omega_{2t}^y. \tag{44}$$

This integral can be simulated by randomly taking $m = 1, \ldots M$ draws from the joint distribution of $\omega$, with draws denoted by $\omega_{mt}$, and determining the fraction of times that the value function for that alternative is the largest among all eight feasible alternatives, that is,

$$\widehat{\Pr}(h_{it} = 0, n_{it} = 0 | \Omega_{it}^{-}) = \frac{1}{M} \sum_{m=1}^{M} I[V_{it}^{00}(\widetilde{\omega_{mt}}) = \max(\widetilde{V_{mt}^{hn}(\Omega_{it}^{-})})], \tag{45}$$

where $I(\cdot)$ is the indicator function equal to one if the statement is true and zero otherwise.

-Similarly form an estimate of the probability for other non-work alternatives, namely for $h_{it} = 0, n_{it} = 1$ for any $t < t_m$ and for $h_{it} = 0$ for any $t_m \leq t \leq T$.

When the wife works, the relevant probability contains the chosen joint alternative $\{h, n\}$ and the observed wage. Consider the case where $h_{it} = 2, n_{it} = 1$ Likelihood contribution for an individual who works 2000 hours in period $t$ at a wage of $w_{it}$ is

$$\Pr(h_{it} = 2, n_{it} = 1, w_{it} | \Omega_{it}^{-}) = \Pr(h_{it} = 2, n_{it} = 1 | w_{it}, \Omega_{it}^{-}) \Pr(w_{it} | \Omega_{it}^{-}) \tag{46}$$

$$= \Pr(w_{it} | \Omega_{it}^{-}) \int_{S_t^{21}(\Omega_{it}^{-})} dF(\omega_{1t}^{1}, \omega_{1t}^{2}, \omega_{1t}^{3}, \omega_{1t}^{N}, \omega_{2t}^{y} | \omega_{2t}^{w}).$$

Suppose that the (log) wage equation is additive in $\eta_{it}^{w}$,

$$\begin{aligned}
\log w_{it} &= \sum_{j=1}^{3} \log r^{j} d_{it}^{j} + \log \Psi_{it}(S_i^w, E_{it}, d_{it-1}^1, d_{it-1}^2, d_{it-1}^3) + \eta_{it}^{w}, \\
&= \sum_{j=1}^{3} \log r^{j} d_{it}^{j} + \log \Psi_{it}(S_i^w, E_{it}, d_{it-1}^1, d_{it-1}^2, d_{it-1}^3) \\
&+ \sum_{m^w=1}^{M} \lambda_{2m^w} 1(type^w = m^w) + \omega_{2it}^{w}
\end{aligned}$$

and further that $\widetilde{\omega}$ is joint normal.

Denoting the deterministic part of the RHS by $\overline{\log w_{it}}$, we can write

$$
\begin{aligned}
\Pr(h_{it} = 2 | w_{it}, \Omega_{it}^{-}) \Pr(w_{it} | \Omega_{it}^{-}) \\
= \int_{S_t^{21}(\Omega_{it}^{-})} dF(\omega_{1t}^{1}, \omega_{1t}^{2}, \omega_{1t}^{3}, \omega_{1t}^{N}, \omega_{2t}^{y} | \omega_{2t}^{w} = \log w_{it} - \overline{\log w_{it}}) \\
\times \frac{1}{w_{it} \sigma_{\omega_2^w}} \phi \left( \frac{\log w_{it} - \overline{\log w_{it}}}{\sigma_{\omega_2^w}} \right)
\end{aligned}
\tag{47}
$$

where $\frac{1}{w_{it}}$ is the Jacobian of the transformation from the distribution of $w$ to the distribution of $\omega_2^w$.

Under our assumptions $f(\omega_{1t}^1, \omega_{1t}^2, \omega_{1t}^3, \omega_{1t}^N, \omega_{2t}^y | \omega_{2t}^w)$ is normal and the frequency simulator for the conditional probability is the same as previously, except that $\omega_{2t}^w$ is set equal to $\log w_{it} - \sum_{j=1}^{3} \log r^j d_{it}^j + \log \Psi_{it} + \sum_{m^w=1}^{M} \lambda_{2m^w} 1(type^w = m^w)$ and the other five $\omega's$ are drawn from $f(\omega_1^1, \omega_1^2, \omega_1^3, \omega_1^N, \omega_2^y | \omega_2^w)$.

Denoting the fixed value of $\omega_{2t}^w$ as $\widehat{\omega_{2t}^w}$,

$$\Pr(h_{it} = 2, n_{it} = 1 | w_{it}, \Omega_{it}^-) \tag{48}$$

$$= \frac{1}{M} \sum_{m=1}^{M} I[V_{it}^{21}(\omega_{m1}^1, \omega_{m1}^2, \omega_{m1}^3, \omega_{m1}^N, \omega_{m2}^y, \widehat{\omega_{mt}^w}) \tag{49}$$

- Although these frequency simulators converge to the true probabilities as $M \to \infty$,there are some practical problems.
- Even for large $M$, the likelihood is not smooth in the parameters, which precludes the use of derivative methods (e.g., BHHH) in maximizing the likelihood and also makes the use of non-derivative methods less efficient.
- Frequency simulators can be smoothed to make the likelihood function differentiable and improve the performance of optimization routines.

# Smoothed logit simulator

- One example is the smoothed logit simulator (McFadden (1989)):

$$\Pr(h_{it}=2, n_{it}=1 | w_{i\tau}, \Omega_{it}^{-}) = \frac{1}{M}\sum_{m=1}^{M} \frac{\exp\left[(V_{itm}^{21} - \max(\widetilde{V_{itm}^{hn}}))/\tau\right]}{\sum_{\{h,n\}} \exp\left[(V_{itm}^{hn} - \max((\widetilde{V_{itm}^{hn}}))/\tau\right]} \tag{50}$$

  where $V_{itm}^{hn}$ is shorthand for the value functions and $\tau$ is a smoothing parameter.
- As $\tau \to 0$, the RHS converges to the frequency simulator.
- The other choice probabilities associated with work alternatives are similarly calculated.

# Alternative Estimation Approaches

- In addition to simulated maximum likelihood, researchers have used varying alternative simulation estimation methods, including minimum distance estimation, simulated method of moments and indirect inference.
- The main limiting factor in estimating DCDP models is the computational burden associated with the iterative process.
- There have been some approaches to reduce the computational burden.

# Hotz and Miller (1993)

- Developed a semi-parametric method for the implementation of DCDP models when errors are additive that does not involve solving the DP model, that is, calculating the $E\max_t$ functions.
- HM prove that the $E\max_t$ functions can be written solely as functions of conditional choice probabilities and state variables for any joint distribution of additive shocks.
- The method doesn't require that errors be extreme value, the computational advantage best exploited with that assumption.

Consider again the binary choice model. If we have an estimate of the conditional choice probabilities at all state points, $E\max_T$ can also be calculated at all state points.

Denoting the (estimate of the) conditional choice probability by $\widehat{\Pr}(d_{iT} = 1|\Omega_{iT}^{-})$,

$$\widehat{E}\max{}_T = \rho\left\{\gamma + \frac{y_T + z\gamma_1 + \gamma_2 h_T - \pi n}{\rho} - \log(\widehat{\Pr}(d_{iT} = 1|\Omega_{iT}^{-})\right\} \tag{51}$$

Consider now period $T-1$ and suppose we have an estimate of the conditional choice probabilities, $\widehat{\Pr}(d_{iT-1}=1|\Omega_{iT-1}^{-})$. Then,

$$E\max_{T-1}=\rho\left\{\gamma+\frac{y_{T-1}+z\gamma_1+\gamma_2 h_{T-1}-\pi n+\delta\widehat{E}\max_T(h_{T-1}+1)}{\rho}-\right.$$

$$\left.\log(\widehat{\Pr}(d_{iT-1}=1|\Omega_{iT-1}^{-})\right\}$$

where, for convenience, we have included only work experience in the $\widehat{E}\max_T$ function.

We can continue substituting the estimated conditional choice probabilities in this recursive manner, yielding at any $t$

$$\widehat{E}\max{}_t = \rho\left\{\gamma + \frac{y_t + z\gamma_1 + \gamma_2 h_t - \pi n + \delta \widehat{E}\max_{t+1}(h_t+1)}{\rho} - \log(\widehat{\Pr}(d_{it} =\right.$$

These $\widehat{E}\max{}_t$ functions can be used in determining the $\xi^*_{it}(\Omega^-_{it})$ cut-off values that enter the likelihood function.

- The empirical strategy involves estimating the conditional choice probabilities from the data (non-parametrically if the data can accommodate it).
- The CCPs correspond to the proportion of women who work for given values of the state variables (for example, work experience).
- Need estimates of the CCPs through the final decision period and for each possible value of the state space.
- Need longitudinal data that either extends to the end of the decision period or assume can be obtained from synthetic cohorts.

- it must also be assumed that there are no state variables observed to the agent but unobserved to us; Arcidiacono and Miller (2007) have developed methods for extending the HM approach to allow for unobserved state variables.
- The convenience additive e.v. errors brings with it the limitations of that assumption previously discussed.
- For policy evaluation, we often need the distributions of the unobservables which are here not estimated.