Dynamic Models
Part 2

Christopher Taber

University of Wisconsin

December 12, 2016

Dynamic Models

We will start with simpler Markov models and then move to Dynamic Discrete Choice Models

I want to define notation to use throughout this set of lecture notes, so I will broadly follow the notation in Arcidiacono and Ellison, “Practical Methods for Estimation of Dynamic Discrete Choice Models” Annual Review of Economics 2011

and in Rust, “Structural Estimation of Markov Decision Processes,” Handbook of Econometrics, 1994

Markov models

In the discrete time duration models there was only one possible state of the world:

- spell underway

and only two possible outcomes

- Spell ends
- Spell continues

Now I want to generalize this to think about more general Markov models

We will assume that you can move into a state of the world d_{it} which potentially takes on multiple (but a discrete number) of outcomes

Examples:

- Working, OLF, Unemployed, In school
- Healthy, Sick, Dead
- Married, Single
- Operate in a market, don't operate in a market
- Have health insurance, don't have health insurance

The main point of this is to get the notation right

Let S_t be the state variables

The outcome variable d_t comes from a set $\mathcal{D}(S_t)$.

Let $\delta^*(S_t)$ map the state variables into the outcome so that

$$d_{it} = \delta_t^*(S_{it}; \theta)$$

where θ is the vector of parameters.

In this case lets write the state variables as consisting of 4 types of variables

$$S_{it} = (d_{t-1}, X_{it}, \mu_i, \varepsilon_{it})$$

where

- d_{t-1} is the current main state we are trying to explain
- X_{it} is observable to the econometrician (and can depend on past values of d_{it})
- μ_i is a vector of unobserved heterogeneity which is not observable to the econometrician and independent of X_{it}
- ε_{it} is a vector of transitory errors that is independent of $X_{it}, \mu_i,$ and $\varepsilon_{i\tau}$ when $\tau \neq t$

If we specify a model for $\delta^* (S_{it}; \theta)$ and the distribution of ε_{it},$F(\varepsilon; \theta)$ then

$$Pr(d_{it} = d \mid d_{it-1}, X_{it}, \mu_i) = \int 1 \left(\delta^* \left(d_{it-1}, X_{it}, \mu_i, \varepsilon; \theta\right) = d\right) dF(\varepsilon; \theta)$$

We also need to specify the evolution of X_{it} which is usually pretty simple

Initial Condition

We need one more part of the model, the initial condition

Start at d_{i0} and assume that d_{i0} is independent of μ_i

Examples

- We are born single and out of work
- A potential firm begins out of the market and decides whether to enter

We can then define the likelihood function as

$$\int \Pi_{t=1}^{T_i} Pr(d_{it} \mid d_{it-1}, X_{it}, \mu) dG(\mu; \theta)$$

where G is the distribution of μ_i

As in the hazard model the initial condition is very important and messy.

Discrete Choice

Before thinking about dynamic discrete choice it makes sense to think about static discrete choice.

Assume that U_{ij} is the utility of individual i at option $j = 0, ..., J$

with

$$U_{ij} = a_j + X_i' \beta_j + Z_j' \delta + v_{ij}$$

(we could have a $Q_{ij}' \gamma_j$ term but lets not worry about that for simplicity)

We assume that there are no ties and that

$$d_i = argmax_{j=\{0,..,J\}} U_{ij}$$

Now think about identification, we get a scale normalization and a location normalization.

It can be seen clearly in the binary choice case $j \in \{0, 1\}$

Choose $j = 1$ if

$$\begin{aligned} & U_{i1} > U_{i0} \\ \Longleftrightarrow & a_1 + X_i'\beta_1 + Z_1'\delta + v_{i1} > a_0 + X_i'\beta_0 + Z_0'\delta + v_{i0} \\ \Longleftrightarrow & \left(a_1 + Z_1'\delta - a_0 - z_0'\delta\right) + X_i'\left(\beta_1 - \beta_0\right) + v_{i1} - v_{i0} > 0 \end{aligned}$$

Clearly all we can identify here is a single intercept

$$a_1 + Z_1'\delta - a_0 - Z_0'\delta$$

(if Z varied across i you could identify δ)

Also can only identify the difference between the betas $(\beta_1 - \beta_0)$ so we can normalize

$$\begin{aligned} a_0 =& 0 \\ \beta_0 =& 0 \\ \delta =& 0 \end{aligned}$$

alternatively we could choose restrict Z to one dimension and estimate the δ on that dimension and then set $a_0 = a_1 = 0$

Error Terms

What about $v_{i1} - v_{i0}$?

Clearly all we can identify is difference

First assume

$$v_i \equiv v_{i1} - v_{i0} \sim N\left(\mu, \sigma^2\right)$$

For the location normalization we can just normalize $\mu = 0$ for the location (or could estimate μ and impose $a_1 = 0$)

So now our model

$$d_i = 1\left(a_1 + X_i'\beta_1 + v_i \geq 0\right)$$

If we multiply a_1, β_1, and v_i by any positive number τ we get exactly the same model

Now we also need a scale normalization

Most common:

- normalize $\sigma = 1$ which gives a probit
- normalize one of the coefficents β_1 to one

Of course there is nothing special about the standard normal and we might want to choose something simpler computationally (there is no closed for solution for the normal cdf)

The other most common assumption is to assume that $v_i = v_{i1} - v_{i0}$ has a logistic distribution for which

$$Pr\left(v_i < \nu\right) = \frac{e^{\nu}}{1 + e^{\nu}}$$

it is also symmetric which means that

$$\begin{aligned} Pr\left(d_i = 1 \mid X_i = x\right) =& Pr\left(a_1 + X_i'\beta_1 \geq -v_i \mid X_i = x\right) \\ =& \frac{e^{a_1 + x'\beta_1}}{1 + e^{a_1 + x'\beta_1}} \end{aligned}$$

the logit model

An alternative assumption gives exactly the same result: suppose that v_{i1} and v_{i0} are independent of each other and both have type I extreme value distribution

$$Pr(v_{ij} \leq \nu) = e^{-e^{-\nu}}$$

then $v_{i1} - v_{i0}$ have a logistic distribution

More than two choices

Now lets go to more than two choices

for simplicity lets focus on 3, but the arguments all apply with more

Now

$$d_i = \begin{cases} 0 & U_{i0} > U_{i1}, U_{i0} > U_{i2} \\ 1 & U_{i1} \geq U_{i0}, U_{i1} > U_{i2} \\ 2 & U_{i2} \geq U_{i0}, U_{i2} \geq U_{i1} \end{cases}$$

so we want to compare

$$\begin{aligned} U_{i0} =& a_0 + X_i'\beta_0 + Z_0'\delta + v_{i0} \\ U_{i1} =& a_1 + X_i'\beta_1 + Z_1'\delta + v_{i1} \\ U_{i2} =& a_2 + X_i'\beta_2 + Z_2'\delta + v_{i2} \end{aligned}$$

We still need a location and scale normalization-but only one

To see location I can subtract U_{i0} from everything without changing the order so that

$$\begin{aligned} U_{i0}^{*} =& 0 \\ U_{i1}^{*} =& (a_1 - a_0) + X_i' (\beta_1 - \beta_0) + (Z_1 - Z_0)' \delta + v_{i1} - v_{i0} \\ U_{i2}^{*} =& (a_2 - a_0) + X_i' (\beta_2 - \beta_0) + (Z_2 - Z_0)' \delta + v_{i2} - v_{i0} \end{aligned}$$

Nothing changes, but I can't subtract anything else so we can use the same normalization

$$\begin{aligned} a_0 =& 0 \\ \beta_0 =& 0 \\ \delta =& 0 \end{aligned}$$

Or we could set $a_1 = a_2 = a_0 = 0$ and estimate a two dimensional δ

Now what about a scale normalization?

Again we only get one:

- If I multiply everything by a postive τ nothing changes
- however, if I mutliply U_{i1} by τ_1 and U_{i2} by $\tau_2 \neq \tau_1$ I change the choice of 2 versus 1

Now with normal error terms if

$$\begin{pmatrix} v_{i1} - v_{i0} \\ v_{i2} - v_{i0} \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}\right)$$

we can do the normalization by setting say $\sigma_{11} = 1$ but still leaving σ_{22} free

Going with more than 3 doesn't fundamentally change things-we still get one location normalization and one scale normalization

Estimation of Multinomial Probit

This is a pain because we have a multiple integral problem, for every option we have another integral

This can become really messy

You also add a ton of new parameters if you have many choices

(There is also an issue about identification-ideally you would have exclusion restrictions)

Multinomial Logit

For computational reasons the multinomial logit is much more popular

If we have J choices and we write

$$U_{ij} = \mu_{ij} + v_{ij}$$

where $\mu_{ij} = a_j + X_i'\beta_j + Z_j'\delta$ after appropriate location and scale normalizations

where the v_{ij} are all independent and type I extreme value we get

$$Pr\left(d_i = j\right) = \frac{e^{\mu_{ij}}}{\sum_{\ell=0}^{J} e^{\mu_{i\ell}}}$$

a closed form answer that is trivial to compute even when J is large

Substitution Patterns

A problem with the multinomial logit is the substitution patterns-you get Independence from Irrelevant Alternatives

The classic example (from McFadden) is if we are looking at transportation choice with three choices

- $j = 0$: Car
- $j = 1$: Red Bus
- $j = 2$: Blue Bus

Think about

$$\frac{Pr\left(d_i = 1\right)}{Pr\left(d_i = 0\right)} = \frac{\frac{e^{\mu_{i1}}}{e^{\mu_{i0}} + e^{\mu_{i1}} + e^{\mu_{i2}}}}{\frac{e^{\mu_{i0}}}{e^{\mu_{i0}} + e^{\mu_{i1}} + e^{\mu_{i2}}}} = \frac{e^{\mu_{i1}}}{e^{\mu_{i0}}}$$

Now suppose we get rid of the Blue bus as an option, now

$$\frac{Pr\left(d_i=1\right)}{Pr\left(d_i=0\right)} = \frac{\frac{e^{\mu_{i1}}}{e^{\mu_{i0}}+e^{\mu_{i1}}}}{\frac{e^{\mu_{i0}}}{e^{\mu_{i0}}+e^{\mu_{i1}}}} = \frac{e^{\mu_{i1}}}{e^{\mu_{i0}}}$$

But this makes no sense-we would expect people who took the blue bus before to substitute towards the red but, but instead they subsitute equally to the red bus and car

This is not just an IO problem

If

- these represent no college, 2 year college and 4 year college
- and we raise the tuition at 4 year college
- we would expect people to substitute more towards 2 year college

To me this while this particular IIA result depends upon the multinomial logit functional form-the more general probelm is assuming that v_{ij} is i.i.d.

We would expect the error term for the blue bus and the error term for the red bus to be highly correlated with eachother.

There are two common solutions to this problem

Nested Logit

The nested logit is one way to get some correlation but still keep things tractable. (more generally using generalized extreme value distribution)

Lets think about a case with one nest.

Partition the choices into L mututally exhaustive categories $C_1, ..., C_L$

We can think of the choice as if it is a two stage process (while it really isn't)

For each ℓ we add a new parameter ρ_ℓ and would choose option

$$Pr\left(d_i = j \mid d_i \in C_\ell\right) = \frac{e^{\mu_{ij}/\rho_\ell}}{\sum_{k \in C_\ell} e^{\mu_{ik}/\rho_\ell}}$$

then we choose the group

$$Pr\left(d_i \in C_\ell\right) = \frac{\left(\sum_{j \in C_\ell} e^{\mu_{ij}/\rho_\ell}\right)^{\rho_\ell}}{\sum_{l=1}^{L}\left(\sum_{j \in C_l} e^{\mu_{ij}/\rho_l}\right)^{\rho_l}}$$

(Note that this is kind of like a nested CES)

Putting them together

$$
\begin{aligned}
Pr\left(d_i = j\right) =& Pr\left(d_i = j \mid d_i \in C_\ell\right) Pr\left(d_i \in C_\ell\right) \\
=& \frac{e^{\mu_{ij}/\rho_\ell}}{\sum_{k \in C_\ell} e^{\mu_{ik}/\rho_\ell}} \frac{\left(\sum_{j \in C_\ell} e^{\mu_{ij}/\rho_\ell}\right)^{\rho_\ell}}{\sum_{l=1}^{L}\left(\sum_{j \in C_l} e^{\mu_{ij}/\rho_l}\right)^{\rho_l}} \\
=& \frac{e^{\mu_{ij}/\rho_\ell}\left(\sum_{j \in C_\ell} e^{\mu_{ij}/\rho_\ell}\right)^{\rho_\ell - 1}}{\sum_{l=1}^{L}\left(\sum_{j \in C_l} e^{\mu_{ij}/\rho_l}\right)^{\rho_l}}
\end{aligned}
$$

Note as well that if all of the $\rho_\ell = 1$ then we are back at the regular multinomial logit

The joint cdf can be written as

$$F_v(\nu) = \exp\left(-\sum_{\ell=1}^{L}\left(\sum_{j\in C_\ell} e^{-\nu_j/\rho_\ell}\right)\right)$$

The correlation of the v_{ij} within a nest is approximately $1-\rho_\ell$ and they are independent across nests

You can also add more nests

Mixed Logit

An alternative way to do this that is quite popular in IO is to used a mixed logit

$$U_{ij} = a_j + X_i'\beta_{ij} + Z_j'\delta_i + v_{ij}$$

particularly with the δ_i.

This makes some real sense as you are allowing people to have preferences for particular aspects of goods

In its simplest form you could just specifiy a distribution for δ_i and integrate through

Goes well beyond this-my IO colleagues have a comparative advantage at teaching this stuff, so I am not going to get into it

OK lets get to dynamics-most of these papers are not going to worry about the substituability issues

Forward Looking Model

Lets combine the discrete choice with the dynamics

Lets start by defining the flow utility for each period as

$$u_t(d_t, X_{it}, \mu_i; \theta) + \varepsilon_{id_t}$$

(starting with linear models for u_t is most common)

I will need to be more explicit about X_{it} at this point-it is observable state variables

- Could include “exogenous variables”
- Could include endogenous variables that depend on previous choices
- Since people are forward looking they will account for this when they make decisions

Let β be the discount rate so now we choose

$$\delta_t^* \left(S_{it};\theta\right) = \underset{d_t \in \mathcal{D}_t(S_{it})}{argmax} E_{i,t,d_t} \left\{ \sum_{\tau=t}^{T} \beta^{\tau-t} (u_\tau(d_\tau, X_{i\tau}, \mu_i;\theta) + \varepsilon_{id_\tau}) \right\}$$

where E_{i,t,d_t} means the expectation of individual i at time t if she chooses option d_t at time t

I will assume the following

- Agents have rational expectations (about future random variables)
- The agent's conditional expectations about X_{it} depend only upon X_{it-1} and d_{it-1}
- Agents also don't have any more information on how ε_{it} will evolve than does the econometrician
- Agents do observe current outcomes of ε_{it} and of μ_i

It is useful to define this using Bellman's equation

Define

$$V_t(S_{it};\theta) \equiv \max_{d_t \in \mathcal{D}_t(S_{it})} E_{i,t,d_t} \left\{ \sum_{\tau=t}^{T} \beta^{\tau-t} \left(u_\tau(d_\tau, X_{i\tau}, \mu_i;\theta) + \varepsilon_{id_\tau} \right) \right\}$$

So we can write

$$V_t(S_{it};\theta) = \max_{d_t \in \mathcal{D}_t(S_{it})} \left\{ u_\tau(d_\tau, X_{i\tau}, \mu_i;\theta) + \varepsilon_{id_\tau} + \beta E_{i,t,d_t} \left[V_{t+1}(S_{it};\theta) \mid X_{it}, d_t, \mu_i \right] \right\}$$

A key result comes from the fact that $\beta E\left[V_{t+1}(S_{it};\theta) \mid X_{it}, d_t, \mu_i\right]$ is only a function of (X_{it}, d_t, μ_i)

That means we can define

$$v_t(d_t, X_{it}, \mu_i; \theta) \equiv u_t(d_t, X_{it}, \mu_i; \theta) + \beta E\left[V_{t+1}(S_{it};\theta) \mid X_{it}, d_t, \mu_i\right]$$

but now we are back in the simpler "static" model

$$\delta_t^*\left(S_{it};\theta\right) = \underset{d_t \in \mathcal{D}_t(S_{it})}{argmax}\left\{v_t(d_t, X_{it}, \mu_i; \theta) + \varepsilon_{id_t}\right\}$$

As long as you can calculate $\beta E\left[V_{t+1}(S_{it};\theta) \mid X_{it}, d_t, \mu_i\right]$ the econometrics is identical to the Markov model

However, this is a big "as long"

How do we usually solve for it?

There are two types of models with two solution methods

- When T is finite we do backward induction.
- When T is infinite we look for a fixed point

I want to focus on the backward induction case

For the infinite case we usually discretize the states

This gives us a finite number of equations and we solve for the fixed point (see Rust)

The ideas are similar, so lets focus on backward induction.

Period T

Start at period T

we can solve for

$$\delta_T^* \left(S_{iT};\theta\right) = \underset{d_T \in \mathcal{D}_T(S_{iT})}{argmax} \left\{u_T(d_T, X_{iT}, \mu_i; \theta) + \varepsilon_{id_T}\right\}$$

Period T-1

Now move to period $T-1$

Let $G(X_{iT} \mid X_{iT-1}, d_{iT-1})$ be the conditional distribution of X_{iT} then

$$E\left[V_T(S_{iT};\theta) \mid X_{iT-1}, d_{T-1}, \mu_i\right] =$$
$$\int\int \left[u_T(\delta_T^*\left(S_{iT};\theta\right), X_T, \mu_i;\theta) + \varepsilon_{i\delta_T^*(S_{iT};\theta)}\right] dF_\varepsilon(\varepsilon) dG(X_T \mid X_{iT-1}, d_{T-1})$$

With some functional form assumptions the integrating over ε can be avoided because closed form solutions are available.

The classic case that gives you a closed form solution is the extreme value.

If all of the ε_{id} are extreme value then we get a really nice result

$$\int \left[u_T(\delta_T^*\left(X_{iT}, \mu_i, \varepsilon_{it}; \theta\right), X_{iT}, \mu_i; \theta) + \varepsilon_{i\delta_T^*(X_{iT}, \mu_i, \varepsilon_{it}; \theta)} \right] dF_\varepsilon(\varepsilon)$$

$$= \log\left(\sum_{d_T \in \mathcal{D}_T(S_{iT})} e^{u_T(d_T, X_{iT}, \mu_i; \theta)} \right) + \gamma$$

where γ is Euler's constant

Another nice example happens with normal error terms and a binary choice variable.

To implement scale and location normalizations assume that

$$\begin{aligned} u_T(1, X_{iT}, \mu_i; \theta) + \varepsilon_{i1} =& u_T(X_{iT}, \mu_i; \theta) + \varepsilon_i \\ u_T(0, X_{iT}, \mu_i; \theta) + \varepsilon_{i0} =& 0 \\ \varepsilon_i =& N(0, \sigma_\varepsilon^2) \end{aligned}$$

Then

$$
\begin{aligned}
&\int \left[u_T(\delta_T^*\left(X_{iT},\mu_i,\varepsilon_{it};\theta\right),X_{iT},\mu_i;\theta)+\varepsilon_{i\delta_T^*(X_{iT},\mu_i,\varepsilon_{it};\theta)}\right]dF_\varepsilon(\varepsilon)\\
=&Pr\left(u_T(X_{iT},\mu_i;\theta)+\varepsilon_i\geq 0\right)E\left(u_T(X_{iT},\mu_i;\theta)+\varepsilon_i \mid u_T(X_{iT},\mu_i;\theta)+\varepsilon_i\geq 0\right)\\
=&\Phi\left(\frac{u_T(X_{iT},\mu_i;\theta)}{\sigma_\varepsilon}\right)u(X_{iT},\mu_i;\theta)+\Phi\left(\frac{u_T(X_{iT},\mu_i;\theta)}{\sigma_\varepsilon}\right)\sigma_\varepsilon\Phi\left(\frac{u_T(X_{iT},\mu_i;\theta)}{\sigma_\varepsilon}\right)\\
=&\sigma_\varepsilon\left[\Phi\left(\frac{u_T(X_{iT},\mu_i;\theta)}{\sigma_\varepsilon}\right)\frac{u_T(X_{iT},\mu_i;\theta)}{\sigma_\varepsilon}+\phi\left(\frac{u_T(X_{iT},\mu_i;\theta)}{\sigma_\varepsilon}\right)\right]
\end{aligned}
$$

However, we still need to worry about the

$$dG(X_T \mid X_{iT}, d_{T-1})$$

part of the expression

This is a mess we have to exactly calculate the value function at all of these points

Often there are a lot of points

Typically people don't do this, they solve at a subset of the points and then use some parametric model to interpolate the other points (see Keane and Wolpin, "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence," 1994)

We have only focused on the last node, but after that we just repeat the exercise

For period $T-1$ we have already calculated $E\left[V_T(S_{iT};\theta) \mid X_{iT-1}, d_{T-1}, \mu_i\right]$ which was the hard part, once we have this we can define

$$
\begin{aligned}
& v_{T-1}(d_{T-1}, X_{iT-1}, \mu_i; \theta) \\
& = u_{T-1}(d_{T-1}, X_{iT-1}, \mu_i; \theta) + \beta E\left[V_T(S_{iT};\theta) \mid X_{iT-1}, d_{T-1}, \mu_i\right]
\end{aligned}
$$

and solve for

$$
\delta^*_{T-1}\left(S_{iT-1};\theta\right) = \underset{d_{T-1}\in\mathcal{D}_T(S_{iT-1})}{argmax} \left\{v_{T-1}(d_{T-1}, X_{iT-1}, \mu_i; \theta) + \varepsilon_{id_{T-1}}\right\}
$$

and then

$$
\begin{aligned}
& E\left[V_{T-1}(S_{iT-1};\theta) \mid X_{iT-2}, d_{T-2}, \mu_i\right] \\
= \int \int & \left[v_{T-1}(\delta^*_{T-1}\left(S_{iT-1},\theta\right), X_{iT-1}, \mu_i;\theta)) + \varepsilon_{i\delta^*_{T-1}(S_{iT-1};\theta)}\right] \\
& dF_\varepsilon(\varepsilon) dG(X_{T-1} \mid X_{iT-2}, d_{T-2})
\end{aligned}
$$

We just keep solving backwards in this way until the initial period

Conditional Choice Probabilities

Hotz and Miller, “Conditional Choice Probabilities and Estimation of Dynamic Models,” REStud, 1993

The idea is more general but the standard case in which it is applied is with extreme value error terms and for now no unobserved heterogeneity

In this case we know that

$$E\left[V_t(S_{it};\theta) \mid X_{it}\right] = \log\left(\sum_{d_t} e^{v_t(d_t,X_{it};\theta)}\right) + \gamma$$

Now take some arbitrary $d_t^* \in \mathcal{D}_t(S_{it})$ from the logit form we know

$$Pr(d_{it} = d_t^* \mid X_{it}) = \frac{e^{v_t(d_t^*,X_{it};\theta)}}{\sum_{d_t} e^{v_t(d_t,X_{it};\theta)}}$$

but then combining these

$$
\begin{aligned}
E\left[V_t(S_{it};\theta) \mid X_{it}\right] =& \log\left(e^{v_t(d_t^*,X_{it};\theta)} \frac{\sum_{d_t} e^{v_t(d_t,X_{it};\theta)}}{e^{v_t(d_t^*,X_{it};\theta)}}\right) + \gamma \\
=& v_t(d_t^*, X_{it};\theta) - \log\left(Pr(d_{it} = d_t^* \mid X_{it})\right) + \gamma
\end{aligned}
$$

But this means that we can write

$$
\begin{aligned}
& v_t(d_{it}, X_{it}; \theta) \\
=& u_t(d_t, X_{it}; \theta) + \beta E\left[V_{t+1}(S_{it}; \theta) \mid X_{it}, d_t\right] \\
=& u_t(d_t, X_{it}; \theta) + \beta \int E\left[V_{t+1}(S_{it+1}; \theta) \mid X_{it+1}\right] dG(X_{it+1} \mid X_{it}, d_t) \\
=& u_t(d_t, X_{it}; \theta) + \beta \int \left[v_{t+1}(d^*_{t+1}, X_{it+1}; \theta) - \log\left(Pr(d_{it+1} = d^*_{t+1} \mid X_{it+1})\right)\right] dG(X_{it+1} \mid X_{it}, d_t) + \beta\gamma
\end{aligned}
$$

Note that we can get $Pr(d_{it+1} = d^*_{t+1} \mid X_{it+1})$ directly from the data

Thus if we knew $v_{t+1}(d^*_{t+1}, X_{it+1}; \theta)$ we wouldn't have to solve the dynamic programming problem

We could just estimate as a nonlinear multinomial logit (and almost linear)

The trick here is to come up with some way to deal with $v_{t+1}(d^*_{t+1}, X_{it+1}; \theta)$

See Arcidiacono, Arcidiacono and Miller, or Hotz and Miller for examples

Hotz and Miller use sterilization as their choice-at that point there are no longer future fertility considerations to be considered so it can be parameterized directly (or normalized to zero)

Unobserved Heterogeneity

The problem here is that this is not implementable when there is unobserved heterogeneity.

Adding it back in gives

$$
\begin{aligned}
& v_t(d_{it}, X_{it}, \mu_i; \theta) \\
=& u_t(d_t, X_{it}, \mu_i; \theta) \\
& + \beta \int \left[v_{t+1}(d^*_{t+1}, X_{it+1}, \mu_i; \theta) - \log \left(Pr(d_{it+1} = d^*_{t+1} \mid X_{it+1}, \mu_i) \right) \right] dG(X_{it+1} \mid X_{it}, d_t) + \beta\gamma
\end{aligned}
$$

But the problem is that $Pr(d_{it+1} = d^*_{t+1} \mid X_{it+1}, \mu_i)$ is not directly identified from the data

This is a big problem in many cases

Arcidicano and Miller, “CCP Estimation of Dynamic Discrete Choice with Unobserved Heterogeneity” (EMA 2011) come up with a solution

First consider the EM Algorithm

EM Algorithm

Think about a case with discrete unobserved heterogeneity so one can write the Log-likelihood function as

$$\sum_i \log\left(\sum_j \pi_j L_j(\theta; Y_i)\right)$$

(One could allow π to depend on X_i but lets focus on the simpler case)

The first order condition with respect to θ is

$$
\begin{aligned}
& \sum_i \frac{\sum_j \pi_j \frac{\partial L_j(\theta; Y_i)}{\partial \theta}}{\sum_j \pi_j L_j(\theta; Y_i)} \\
= & \sum_i \sum_j \left[\frac{\pi_j}{\sum_j \pi_j L_j(\theta; Y_i)} \right] \frac{\partial L_j(\theta; Y_i)}{\partial \theta} \\
= & \sum_i \sum_j \left[\frac{\pi_j L_j(\theta; Y_i)}{\sum_j \pi_j L_j(\theta; Y_i)} \right] \frac{\partial \log\left(L_j(\theta; Y_i)\right)}{\partial \theta} \\
\equiv & \sum_i \sum_j q(j \mid Y_i; \theta) \frac{\partial \log\left(L_j(\theta; Y_i)\right)}{\partial \theta}
\end{aligned}
$$

where from Bayes theorem, q is the conditional probability that the unobservable is node j conditional on the data and parameter vector θ

But that means by the law of iterated expectations

$$\pi_j \approx \frac{1}{N} \sum_i q(j \mid Y_i; \theta)$$

This suggests a two staged process

M Stage

In the M (Maximization) phase we take $\widehat{q}(j \mid Y_i; \theta)$ as given (from previous stage) and maximize

$$\sum_i \sum_j \widehat{q}(j \mid Y_i; \theta) \log \left(L_j(\theta; Y_i)\right)$$

to get an estimate $\widehat{\theta}$

E Stage

In E (Expectation) phase we take parameters $\widehat{\theta}$ and $\widehat{\pi}$ as given (from previous stage) and calculate

$$\widehat{q}(j \mid Y_i; \widehat{\theta}) = \frac{\widehat{\pi}_j L_j(\widehat{\theta}; Y_i)}{\sum_j \widehat{\pi}_j L_j(\widehat{\theta}; Y_i)}$$

and that will also yield a new

$$\widehat{\pi}_j \approx \frac{1}{N} \sum_i \widehat{q}(j \mid Y_i; \widehat{\theta})$$

We keep iterating until we find a fixed point

The point it converges to will be a point that solves the first order condition of the MLE problem

This is generally **not** computationally better than MLE because we may need to solve the maximization step many times

However, solving the M- step might be much easier than the full model

Arcidiacono and Jones, EMA, 2003 give some examples of these cases

CCP and the EM Algorithm

CCP is another case like that

Recall that the problem with unobserved heterogeneity was that we couldn't observe $Pr(d_{it} = d_t^* \mid X_{it}, \mu_i)$ in the data

Using a bit odd notation we could think of μ_i taking on j values, $j = 1, ..., K$ then notice that

$$
\begin{aligned}
& Pr(d_{it} = d_t^* \mid X_{it} = x, \mu_i = j]) \\
& \quad = \frac{E(1(d_{it} = d_t^*)\, 1(\mu_i = j) \mid X_{it} = x)}{E(1(\mu_i = j) \mid X_{it} = x)} \\
& \quad = \frac{E(E(1(d_{it} = d_t^*)\, 1(\mu_i = j) \mid Y_i) \mid X_{it} = x)}{E(E(1(\mu_i = j) \mid Y_i) \mid X_{it} = x)} \\
& \quad = \frac{E(1(d_{it} = d_t^*)\, q(j \mid Y_i; \theta) \mid X_{it} = x)}{E(q(j \mid Y_i; \theta) \mid X_{it} = x)}
\end{aligned}
$$

(Y_i is the full set of observables so it includes d_{it})

Notice that we can approximate this as

$$\frac{\sum_i 1\left(d_{it} = d_t^*\right) q(j \mid Y_i; \theta) 1\left(X_{it} = x\right)}{\sum_i q(j \mid Y_i; \theta) 1\left(X_{it} = x\right)}$$

Arcidiacono and Miller show that you can

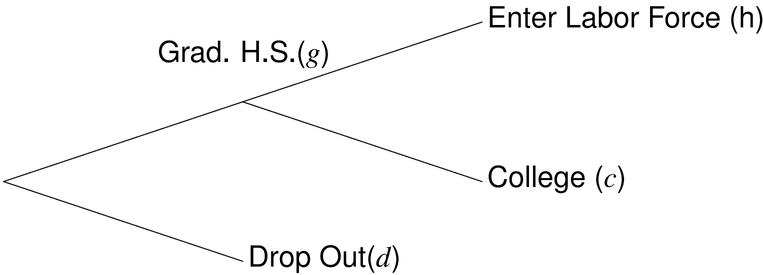
- Calculate this in the E step
- Use this estimate for the CCP in the M stage

Since you don't need to solve the dynamic programming model this can be really quick

Identification

Taber , "Semiparametric Identification and Heterogeneity in Dynamic Discrete Choice Models," Journal of Econometrics, 2000.

I use the simplest Dynamic model you can think about using education as an example (like the Cameron and Heckman case)


Enter Labor Force (h)
Grad. H.S.(g)
College (c)
Drop Out(d)

Define the model in terms of lifetime utility at the terminal nodes

With an exclusion restriction model is

$$\begin{aligned} V_{ci} &= g_c(X_{ci}, X_{0i}) + \varepsilon_{ci} \\ V_{di} &= g_d(X_{di}, X_{0i}) + \varepsilon_{di} \\ V_{hi} &= 0 \end{aligned}$$

and $J_i \in \{c, d, h\}$ the option that individual i actually chose

The key complication is exactly what the agent knows at time 1 about the error term at time 2

Let this be summarized by ε_{1i}

We let X_{1i} denote other information the agent might have about future values of X_i

I use the following timing

Known to the Agent at time one	Learned by the Agent at time two	Observed by the Econometrician
$\varepsilon_{1i}, \varepsilon_{di}$	ε_{ci}	X_{0i}, X_{1i}, X_{di}
X_{0i}, X_{1i}, X_{di}	X_{ci}	X_{ci}
$G(\cdot)$		J_i

Solving the model backward the person attends college if

$$V_{ci} > V_{hi} \Longleftrightarrow g_c(X_{ci}, X_{0i}) + \varepsilon_{ci} > 0$$

The only place where dynamics is interesting is the g node

I define

$$\begin{aligned} & V_g(x_1, x_d, x_0, \epsilon_1) \\ & \equiv E[\max\{V_{ci}, V_{hi}\} \mid (X_{1i}, X_{di}, X_{0i}) = (x_1, x_d, x_0), \varepsilon_{1i} = \epsilon_1] \end{aligned}$$

This person will graduate from college when

$$V_g(X_{1i}, X_{di}, X_{0i}, \varepsilon_{1i}) > V_{hi}$$

Identification of g_c is like the standard “identification at infinity” argument for any selection model

$$\lim_{g_d(x_d,x_0)\to-\infty} \Pr(J_i = c \mid X_i = x)$$
$$= \lim_{g_d(x_d,x_0)\to-\infty} \Pr[g_d(x_d,x_0) + \varepsilon_{di} \le V_g(x_1,x_d,x_0,\varepsilon_{1i}), g_c(x_c,x_0) + \varepsilon_{ci} > 0]$$
$$= \Pr[g_c(x_c,x_0) + \varepsilon_{ci} > 0].$$

Identifying g_d is kind of similar, suppose that you have an X_{1i} such that when it goes to $-\infty$ the distribution of g_c shifts to the left.

(this is easiest to think about when X_{ci} is known at time 1 so that $X_{ci} = X_{1i}$)

Then

$$\lim_{x_1 \to -\infty} E\left[\max\left(g_c(X_{ci}, x_0) + \varepsilon_{ci}, 0\right) \mid (X_{1i}, X_{di}, X_{0i}) = (x_1, x_d, x_0), \varepsilon_{1i} = \epsilon_1\right]$$
$$= 0,$$

so that,

$$\begin{aligned}
&\lim_{x_1 \to -\infty} \Pr(J_i = d \mid X_i = x) \\
&= \lim_{x_1 \to -\infty} \Pr[g_d(x_d, x_0) + \varepsilon_{di} > E\left[\max\left(V_{ci}, 0\right) \mid (x_1, x_d, x_0), \epsilon_1\right]] \\
&= \Pr[g_d(x_d, x_0) + \varepsilon_{di} > 0].
\end{aligned}$$

The error terms are not identified without putting more structure on ε_{1i}

I cover two cases:

1. $\varepsilon_{1i} = \varepsilon_{di}$
2. $\varepsilon_{ci} = \mu_{ci} + \eta_{ci}$ where μ_{ci} is known at time 1 and η_{ci} is independent of anything known at time 1

We know that

$$Pr(J_i = h \mid X_i) = Pr(g_{di} + \varepsilon_{di} \leq V_g(X_{1i}, \varepsilon_{1i}), g_{ci} + \varepsilon_{ci} > 0).$$

So by sending $V_g(X_1, \varepsilon_1) \to 0$ I can identify

$$Pr(g_{di} + \varepsilon_{di} \leq 0, g_{ci} + \varepsilon_{ci} > 0).$$

but given that g_a and g_b are identified, I can identify the joint distribution of $(\varepsilon_{di}, \varepsilon_{ci})$

- For the first model this is everything so we are done
- The second takes slightly more work
  - choose X_{1i} and X_{di} so that:
    - $Prob(J_i = c \mid X_{1i}) \to 1$
    - That implies that $V_g(X_{1i}, \varepsilon_{1i}) \to E(g_{ci} \mid X_{1i}) + \mu_{ci}$
    - which unfortunately also implies that $E(g_{ci} \mid X_{1i}) \to \infty$
    - So we need to also send $g_{di} \to \infty$ at the same rate so that $g_{di} - E(g_{ci} \mid X_{1i}) = \widetilde{g}_i$
  - OK thats a mess, but once we do that we can identify
  - $$Pr(J_i = c \mid X_i) \approx Pr(\widetilde{g}_i + \varepsilon_{di} > \mu_{ci}, g_{ci} + \mu_{ci} + \eta_{ci})$$
  - It turns out that knowledge of the marginal distribution of ε_{di} and the joint distribution of $(\varepsilon_{di} - \mu_{ci}, \mu_{ci} + \eta_{ci})$ is enough to identify the joint distribution of $(\varepsilon_{di}, \mu_{ci})$ and of η_{ci} (using characteristic functions)

Examples

Lets look at probably two most classic examples

Harold Zurcher

Rust, “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher,”EMA, 1987

Harold Zurcher managed the bus depot here in Madison

TABLE I

Bus Types Included in Sample

Bus Group	Number of Buses	Manufacturer	Engine	Model	Year	Seats	Empty Weight	Purchase Price	Estimated Value as of 10/1/84
1	15	Grumman	V6-92 series	870	1983	48	25,800	$145,097	$145,097
2	4	Chance	3208 CAT	RT-50	1981	10*	N.A.	100,775	124,772
3	48	GMC	8V71	T8H203	1979	45	25,027	92,668	125,000
4	37	GMC	8V71	5308A	1975	53	20,955	62,506	55,000
5	12	GMC	8V71	5308A	1974	53	20,955	49,975	48,000
6	10	GMC	6V71	4523A	1974	45	19,274	45,704	48,000
7	18	GMC	8V71	5308A	1972	51	20,955	43,856	45,000
8	18	GMC	6V71	4523A	1972	45	19,274	40,542	40,000

His choice each period is whether to

- Do routine maintenance on the bus $d_{it} = 0$
- Completely rebuild the engine $d_{it} = 1$ which makes it like new

There are two state variables

- Mileage on engine X_t which is observable-evolves according to a distribution of $X_{t+1} - X_t$ which is i.i.d. $g(\cdot; \theta_3)$
- Unobservable extreme value terms ε_{di} which are i.i.d. extreme value

$$q(\varepsilon; \theta_2) = e^{-\varepsilon+\theta_2} e^{-e^{-\varepsilon+\theta_2}}$$

with θ_2 as Euler's constant which gives multinomial logit

TABLE IIa

Summary of Replacement Data

(Subsample of buses for which at least 1 replacement occurred)

	Mileage at Replacement				Elapsed Time (Months)				
Bus Group	Max	Min	Mean	Standard Deviation	Max	Min	Mean	Standard Deviation	Number of Observations
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	273,400	124,800	199,733	37,459	74	38	59.1	10.9	27
4	387,300	121,300	257,336	65,477	116	28	73.7	23.3	33
5	322,500	118,000	245,291	60,258	127	31	85.4	29.7	11
6	237,200	82,400	150,786	61,007	127	49	74.7	35.2	7
7	331,800	121,000	208,963	48,981	104	41	68.3	16.9	27
8	297,500	132,000	186,700	43,956	104	36	58.4	22.2	19
Full Sample	387,400	83,400	216,354	60,475	127	28	68.1	22.4	124

TABLE IIb

CENSORED DATA

(Subsample of buses for which no replacements occurred)

	Mileage at May 1, 1985				Elapsed Time (months)				
Bus Group	Max	Min	Mean	Standard Deviation	Max	Min	Mean	Standard Deviation	Number of Observations
1	120,151	65,643	100,117	12,929	25	25	25	0	15
2	161,748	142,009	151,183	8,530	49	49	49	0	4
3	280,802	199,626	250,766	21,325	75	75	75	0	21
4	352,450	310,910	337,222	17,802	118	117	117.8	0.45	5
5	326,843	326,843	326,843	0	130	130	130	0	1
6	299,040	232,395	265,264	33,332	130	128	129.3	1.15	3
7	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0
Full Sample	352,450	65,643	207,782	85,208	130	25	66.4	34.6	49

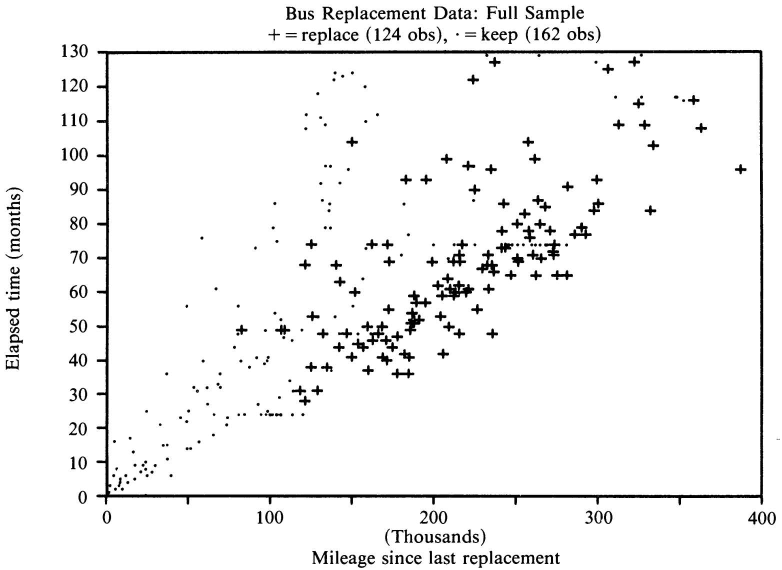

Bus Replacement Data: Full Sample
+ = replace (124 obs), · = keep (162 obs)
Elapsed time (months)
0
10
20
30
40
50
60
70
80
90
100
110
120
130
0
100
200
300
400
(Thousands)
Mileage since last replacement

FIGURE 1

The utility is modeled as

$$u(0, X_{it}; \theta) = -c(X_{it}; \theta_1)$$
$$u(1, X_{it}; \theta) = RC - c(0; \theta_1)$$

There are no t subscripts because this is a stationary infinitely lived problem.

The problem is solved using a nested fixed point algorithm (see Rust for details)

The parameters are

$$\theta = \{\beta, \theta_1, RC, \theta_3\}$$

Take β (and θ_2) as given

X_{it} is discretized into intervals of 5000

Distribution of ΔX_{it} is divided into 3 categories

- 0-5000
- 5000-10,000
- 10,000-∞

Thus there are only 2 parameters in θ_{30} and θ_{31} to cover these three probabilities

He tries different specifications for the cost functions-linear works fine

He estimates the model using full maximum likelihood

TABLE IX

STRUCTURAL ESTIMATES FOR COST FUNCTION $c(x, \theta_1) = .001\theta_{11}x$
FIXED POINT DIMENSION = 90
(Standard errors in parentheses)

Parameter		Data Sample			Heterogeneity Test	
Discount Factor	Estimates/ Log-Likelihood	Groups 1, 2, 3 3864 Observations	Group 4 4292 Observations	Groups 1, 2, 3, 4 8156 Observations	LR Statistic ($df = 4$)	Marginal Significance Level
$\beta = .9999$	RC	11.7270 (2.602)	10.0750 (1.582)	9.7558 (1.227)	85.46	1.2E−17
	θ_{11}	4.8259 (1.792)	2.2930 (0.639)	2.6275 (0.618)		
	θ_{30}	.3010 (.0074)	.3919 (.0075)	.3489 (.0052)		
	θ_{31}	.6884 (.0075)	.5953 (.0075)	.6394 (.0053)		
	LL	−2708.366	−3304.155	−6055.250		
$\beta = 0$	RC	8.2985 (1.0417)	7.6358 (0.7197)	7.3055 (0.5067)	89.73	1.5E−18
	θ_{11}	109.9031 (26.163)	71.5133 (13.778)	70.2769 (10.750)		
	θ_{30}	.3010 (.0074)	.3919 (.0075)	.3488 (.0052)		
	θ_{31}	.6884 (.0075)	.5953 (.0075)	.6394 (.0053)		
	LL	−2710.746	−3306.028	−6061.641		
Myopia test:	LR Statistic ($df = 1$)	4.760	3.746	12.782		
$\beta = 0$ vs. $\beta = .9999$	Marginal Significance Level	0.0292	0.0529	0.0035		

Keane and Wolpin

Keane and Wolpin, “The Career Decisions of Young Men," JPE, 1997 This is the best known structural model of labor dynamics

There have been many subsequent papers written that use the basic framework, but build on it

I discuss the first classic paper

Essentially a dynamic Roy model

Basic Model

People start making decisions at age $a = 16$ and live until age A.

At each age they can choose one of 5 options:

1. Work in Blue Collar Job
2. Work in White Collar Job
3. Work in Military
4. Go to School
5. Home Production

For each of these 5 options let:

- $d_m(a)$ be an indicator for whether option m was chosen
- $R_m(a)$ be the conditional reward if m was chosen
- $g(a)$ schooling at age a

Then

$$
\begin{aligned}
R(a) &= \sum_{m=1}^{5} R_m(a) d_m(a) \\
g(a) &= \sum_{\alpha=1}^{a-1} d_4(\alpha)
\end{aligned}
$$

Consider each of the three working options ($m = 1, 2, 3$) then let

- $e_m(a)$ skill level in occupation m
- r_m rental rate in occupation m
- $x_m(a)$ work experience in occupation m,($x_m(a) = \sum_{\alpha=1}^{a-1} d_m(\alpha)$)

They assume that

$$log(e_m(a)) = e_m(16) + e_{m1}g(a) + e_{m2}x_m(a) - e_{m3}x_m^2(a) + \varepsilon_m(a)$$

for $m = 1, 2, 3$, and $a = 1, ..., A$.

Since people only care about wages (no hours dimension of labor supply or tastes)

$$
\begin{aligned}
R_m(a) &= w_m(a) \\
&= r_m exp(e_m(16) + e_{m1} g(a) + e_{m2} x_m(a) - e_{m3} x_m^2(a) + \varepsilon_m(a))
\end{aligned}
$$

They define the reward functions for the other two alternatives as:

$$R_4(a) = e_4(16) - tc_1 1\left[g(a) \geq 12\right] - tc_2 1\left[g(a) \geq 16\right] + \varepsilon_4(a)$$
$$R_5(a) = e_5(16) + \varepsilon_5(a)$$

and further define:

$$\begin{aligned}
\varepsilon(a) &\equiv \{\varepsilon_1(a), \varepsilon_2(a), \varepsilon_{13}(a), \varepsilon_4(a), \varepsilon_5(a)\} \sim N(0, \Omega) \\
e(16) &\equiv \{e_1(16), e_2(16), e_3(16), e_4(16), e_5(16)\} \\
x(a) &\equiv \{x_1(a), x_2(a), x_3(a))\} \\
S(a) &\equiv \{e(16), g(a), x(a), \varepsilon(a))\}
\end{aligned}$$

We are done with the model, the agents just solve the dynamic programming problem

$$V(S(a), a) =$$
$$max_m \left[R_m(S(a), a) + \delta E(V(S(a+1), a+1) \mid S(a), d_m(a) = 1)\right]$$

for $a < A$

In the last period

$$V(S(A), A) = max_m \left[R_m(S(A), A)\right]$$

Thats it, that is the whole model.

They solve backward interpolating between different points in the state space

Estimation

Keane and Wolpin use the NLSY79 data set, starting with people age 16 who they observe until a certain age (call it $\bar{a}_i$ for individual i).

They also observe schooling $(g_i(a))$, sector specific experience $(x_i(a))$, and choices made at each age until $\bar{a}_i$.

They will allow for heterogeneity in $\varepsilon_i(a)$ which is unobservable

They also will allow for heterogeneity in initial endowments as well $e_i(16)$ although this is not observable to the econometrician.

Given the model it is straight forward (though computationally intensive) to calculate

$$Pr(c_i(a) \mid a, g_i(a), x_i(a), e_i(16); \theta)$$

with knowledge of the other parameters θ.

Thus if we know $e_i(16)$ the likelihood for individual i would be straight forward to calculate because there is no serial correlation in $\varepsilon_i(a)$.

$$\mathcal{L}_i(e_i(16), \theta) \equiv \prod_{a=16}^{\bar{a}_i} Pr(c_i(a) \mid a, g_i(a), x_i(a), e_i(16); \theta)$$

To deal with heterogeneity they assume that there are a finite number of types (Heckman/Singer style)

Assume that there are K types and let π_k denote the proportion in the population of type k

further let $e^k(16)$ denote the vector of skills for type k

Then the likelihood takes the form:

$$\mathcal{L}_i(\theta, \pi, e(16)) = \sum_{k=1}^{K} \mathcal{L}_i(e^k(16), \theta)\pi_k$$

Thats the model, now it is just time to calculate it.

TABLE 1

CHOICE DISTRIBUTION: WHITE MALES AGED 16–26

AGE	School	Home	White-Collar	Blue-Collar	Military	TOTAL
16	1,178	145	4	45	1	1,373
	85.8	10.6	.3	3.3	.1	100.0
17	1,014	197	15	113	20	1,359
	74.6	14.5	1.1	8.3	1.5	100.0
18	561	296	92	331	70	1,350
	41.6	21.9	6.8	24.5	5.2	100.0
19	420	293	115	406	107	1,341
	31.3	21.9	8.6	30.3	8.0	100.0
20	341	273	149	454	113	1,330
	25.6	20.5	11.2	34.1	8.5	100.0
21	275	257	170	498	106	1,306
	21.1	19.7	13.0	38.1	8.1	100.0
22	169	212	256	559	90	1,286
	13.1	16.5	19.9	43.5	7.0	100.0
23	105	185	336	546	68	1,240
	8.5	14.9	27.1	44.0	5.5	100.0
24	65	112	284	416	44	921
	7.1	12.2	30.8	45.2	4.8	100.0
25	24	61	215	267	24	591
	4.1	10.3	36.4	45.2	4.1	100.0
26	13	32	88	127	2	262
	5.0	12.2	33.6	48.5	.81	100.0
Total	4,165	2,063	1,724	3,762	645	12,359
	33.7	16.7	14.0	30.4	5.2	100.0

NOTE.—Number of observations and percentages.

TABLE 2

TRANSITION MATRIX: WHITE MALES AGED 16–26

	CHOICE (t)				
CHOICE ($t-1$)	School	Home	White-Collar	Blue-Collar	Military
School:					
Row %	69.9	12.4	6.5	9.9	1.3
Column %	91.2	32.6	2.5	14.2	11.2
Home:					
Row %	9.8	47.2	8.1	31.3	3.7
Column %	4.4	42.9	8.8	15.6	10.7
White-collar:					
Row %	5.7	6.3	67.4	19.9	.7
Column %	1.8	4.0	51.4	7.0	1.4
Blue-collar:					
Row %	3.4	12.4	9.9	73.4	.9
Column %	2.6	19.0	18.2	61.7	4.3
Military:					
Row %	1.4	5.5	3.1	9.6	80.5
Column %	.2	1.6	1.0	1.5	72.4

TABLE 3

SELECTED CHOICE-STATE COMBINATIONS

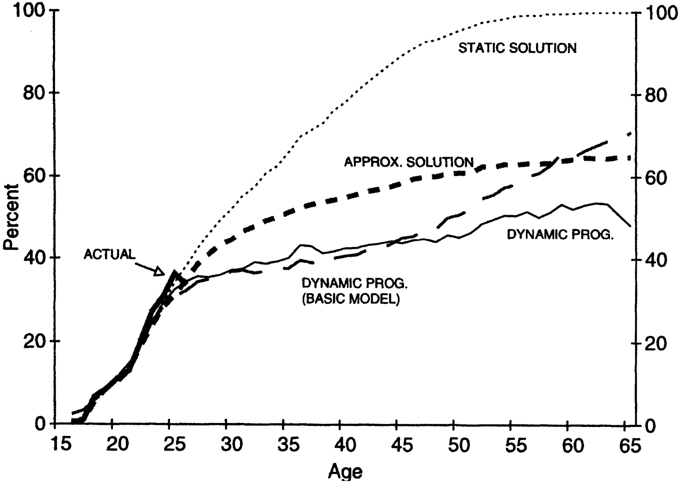
Highest grade completed	9	10	11	12	13	14	15	16	17
Percentage choosing school	26.9	59.8	49.1	13.5	45.1	44.8	62.5	13.5	42.5
If in school previous period	73.5	91.1	85.0	44.2	72.9	70.6	68.8	23.5	55.6
White-collar experience	0	1	2	3	4	5	6		
Percentage choosing white-collar employment	6.8	38.0	55.3	63.3	76.2	74.6	79.2		
If white-collar previous period	...	57.5	71.7	76.7	78.8	82.0	86.4		
Blue-collar experience	0	1	2	3	4	5	6	7	
Percentage choosing blue-collar employment	15.0	51.6	64.9	74.0	74.9	81.2	77.1	88.3	
If blue-collar previous period	...	62.0	71.4	78.7	81.7	85.3	78.7	85.4	
Military experience	0	1	2	3	4	5			
Percentage choosing military employment	1.5	68.0	56.6	44.6	32.7	61.9			
If military previous period	...	90.7	86.5	74.0	57.1	78.8			

TABLE 4

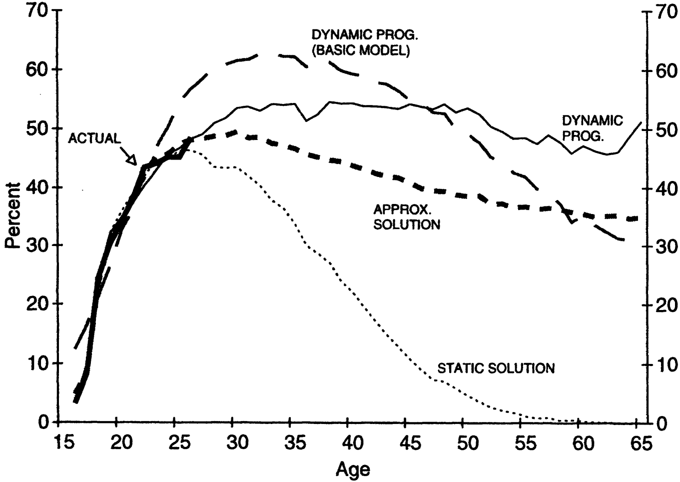
AVERAGE REAL WAGES BY OCCUPATION: WHITE MALES AGED 16–26

	Mean Wage			
Age	All Occupations	White-Collar	Blue-Collar	Military
16	10,217 (28)	. . .	10,286 (26)	. . .
17	11,036 (102)	10,049 (14)	11,572 (75)	9,005 (13)
18	12,060 (377)	11,775 (71)	12,603 (246)	10,171 (60)
19	12,246 (507)	12,376 (97)	12,949 (317)	9,714 (93)
20	13,635 (587)	13,824 (128)	14,363 (357)	10,852 (102)
21	14,977 (657)	15,578 (142)	15,313 (419)	12,619 (96)
22	17,561 (764)	20,236 (214)	16,947 (476)	13,771 (74)
23	18,719 (833)	20,745 (299)	17,884 (481)	14,868 (53)
24	20,942 (667)	24,066 (259)	19,245 (373)	15,910 (35)
25	22,754 (479)	24,899 (207)	21,473 (250)	17,134 (22)
26	25,390 (206)	32,756 (79)	20,738 (125)	. . .

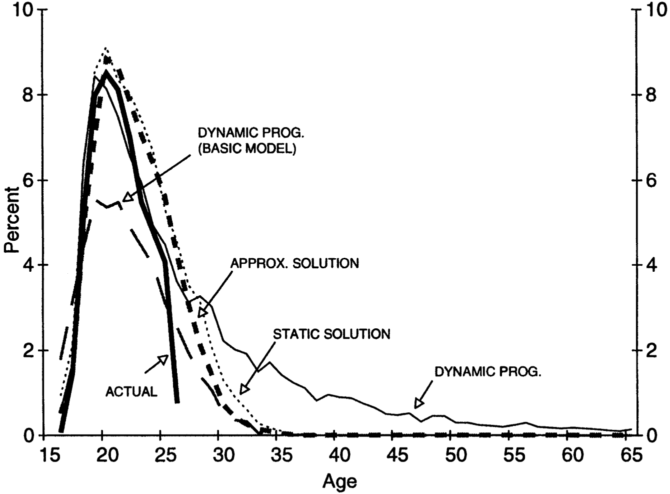
NOTE.—Number of observations is in parentheses. Not reported if fewer than 10 observations.



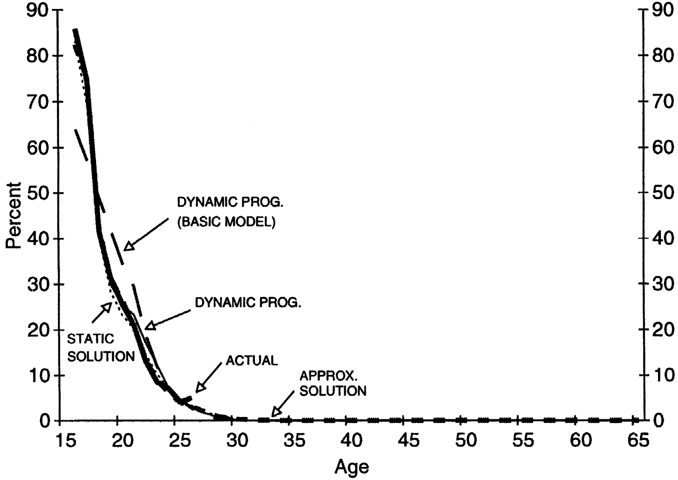
100
80
60
40
20
0
Percent
15
20
25
30
35
40
45
50
55
60
65
Age
STATIC SOLUTION
APPROX. SOLUTION
ACTUAL
DYNAMIC PROG.
DYNAMIC PROG.
(BASIC MODEL)



DYNAMIC PROG.
(BASIC MODEL)
ACTUAL
DYNAMIC
PROG.
APPROX.
SOLUTION
STATIC SOLUTION
Percent
Age
70
60
50
40
30
20
10
0
15
20
25
30
35
40
45
50
55
60
65



DYNAMIC PROG.
(BASIC MODEL)
APPROX. SOLUTION
STATIC SOLUTION
DYNAMIC PROG.
ACTUAL
Percent
Age



90
80
70
60
50
40
30
20
10
0
Percent
15
20
25
30
35
40
45
50
55
60
65
Age
DYNAMIC PROG.
(BASIC MODEL)
DYNAMIC PROG.
STATIC
SOLUTION
ACTUAL
APPROX.
SOLUTION

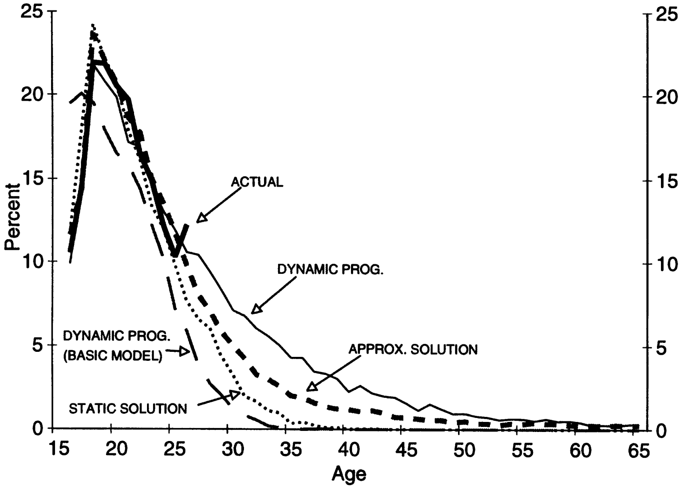


FIG. 5.—Percentage at home by age

TABLE 5

χ^2 GOODNESS-OF-FIT TESTS OF THE WITHIN-SAMPLE CHOICE DISTRIBUTION: DYNAMIC PROGRAMMING MODEL AND MULTINOMIAL PROBIT

Age	School	Home	White-Collar	Blue-Collar	Military	Row
16:						
DP-basic	103.05*	17.10*	†	92.61*	†	213.2*
DP-extended	.00	.07	†	.15	†	.22
APP	2.00	.19	†	7.05*	†	9.24*
17:						
DP-basic	74.13*	7.37*	21.14*	54.63*	11.86*	169.15*
DP-extended	.95	.02	.28	3.31	.42	4.98
APP	.02	.00	1.78	.03	.00	1.84
18:						
DP-basic	15.02*	1.60	2.18	6.75*	1.71	27.26*
DP-extended	.03	.00	.93	.01	3.09	4.06
APP	.09	.94	3.03	.42	.17	4.65
19:						
DP-basic	35.83*	5.04*	.26	7.23*	14.41*	62.77*
DP-extended	.83	.51	.07	1.27	.34	3.02
APP	.00	.02	.01	.17	1.53	1.73
20:						
DP-basic	31.10*	6.24*	.14	.92	24.47*	62.86*
DP-extended	.16	.25	.24	.22	.22	.94
APP	.25	.01	.82	.06	.17	1.31
21:						
DP-basic	31.28*	6.54*	.01	1.46	16.61*	55.89*
DP-extended	2.91	3.50	2.45	.23	.72	9.81*
APP	.00	.65	.05	.03	.41	1.14
22:						
DP-basic	23.78*	2.94	1.01	.08	11.84*	39.66*
DP-extended	12.43*	.11	.61	3.04	.38	16.60*
APP	.12	1.49	.72	.64	1.21	4.19
23:						
DP-basic	12.63*	7.78*	2.99	2.00	3.15	28.56*
DP-extended	14.66*	.12	3.76	.42	.44	19.40*
APP	.23	.14	5.90*	.44	4.38	10.97*
24:						
DP-basic	.18	4.76*	2.28	4.61*	1.40	13.30*
DP-extended	.18	.99	.81	.04	.04	1.89
APP	1.21	2.77	2.20	.05	2.77	10.01*
25:						
DP-basic	.30	12.35*	6.21*	9.31*	1.84	30.01*
DP-extended	.14	3.45	2.71	.29	.23	6.82
APP	.01	2.98	5.00*	.61	2.56	11.16*
26:						
DP-basic	4.96*	38.64*	.17	3.13	†	46.90*
DP-extended	2.61	2.14	.45	.00	†	5.20
APP	2.84	4.95*	.10	.01	†	7.90*

NOTE.—The basic dynamic programming (DP-basic) model has 50 parameters, the extended dynamic programming (DP-extended) model has 83 parameters, and the approximate decision rule (APP) model has 75 parameters.

* Statistically significant at the .05 level.

† Fewer than five observations.

TABLE 6

WITHIN-SAMPLE WAGE FIT

	WHITE-COLLAR				BLUE-COLLAR			
	NLSY*	DP-Basic	DP-Extended	Static	NLSY†	DP-Basic	DP-Extended	Static
Wage:								
Mean	19,691	17,456	19,605	19,688	16,224	16,230	15,805	15,914
Standard deviation	12,461	10,324	12,091	13,664	8,631	8,437	8,431	9,837
Wage regression:								
Highest grade completed	.095	.033	.090	.091	.048	.006	.047	.056
	(.007)‡	(.007)	(.006)	(.007)	(.008)	(.006)	(.006)	(.007)
Occupation-specific experience	.103	.017	.080	.123	.096	.082	.078	.108
	(.009)	(.011)	(.012)	(.010)	(.005)	(.004)	(.004)	(.005)
Constant	8.33	9.15	8.44	8.22	8.80	9.25	8.84	8.54
	(.102)	(.087)	(.080)	(.100)	(.096)	(.069)	(.078)	(.082)
R^2	.213	.021	.182	.172	.150	.117	.104	.142
Observations	1,509	1,605	1,685	1,698	3,143	4,013	3,761	3,772

* Three wage outliers of over $250,000 were discarded. The only important effect was to reduce the wage standard deviation significantly.
† Two wage outliers of over $200,000 were discarded. The only important effect was to reduce the wage standard deviation significantly.
‡ Heteroskedasticity-corrected standard errors are in parentheses.

Given that the model does not fit that well, Keane and Wolpin do several things to improve the fit of the model:

1. More terms are added to the civilian wage equations
2. Allow for a reward cost if you switch occupations, and larger if you start a new occupation
3. Include non-wage tastes for the occupations
4. Include a consumption value of school, a cost of reentry to school, and a psychic cost of getting high school/college diploma
5. Payoff for home production change by age

Here are the results

TABLE 7

Estimated Occupation-Specific Parameters

	White-Collar	Blue-Collar	Military
	1. Skill Functions		
Schooling	.0700 (.0018)	.0240 (.0019)	.0582 (.0039)
High school graduate	−.0036 (.0054)	.0058 (.0054)	...
College graduate	.0023 (.0052)	.0058 (.0080)	...
White-collar experience	.0270 (.0012)	.0191 (.0008)	...
Blue-collar experience	.0225 (.0008)	.0464 (.0005)	...
Military experience	.0131 (.0023)	.0174 (.0022)	.0454 (.0037)
"Own" experience squared/100	−.0429 (.0032)	−.0759 (.0025)	−.0479 (.0140)
"Own" experience positive	.1885 (.0132)	.2020 (.0128)	.0753 (.0344)
Previous period same occupation	.3054 (.1064)	.0964 (.0124)	...
Age*	.0102 (.0005)	.0114 (.0004)	.0106 (.0022)
Age less than 18	−.1500 (.0515)	−.1433 (.0308)	−.2539 (.0443)
Constants:			
Type 1	8.9370 (.0152)	8.8811 (.0093)	8.540 (.0234)
Deviation of type 2 from type 1	−.0872 (.0089)	.3050 (.0138)	...
Deviation of type 3 from type 1	−.6091 (.0143)	−.2118 (.0144)	...
Deviation of type 4 from type 1	−.5200 (.0199)	−.0547 (.0177)	...
True error standard deviation	.3864 (.0094)	.3823 (.0074)	.2426 (.0249)
Measurement error standard deviation	.2415 (.0140)	.1942 (.0134)	.2063 (.0207)
Error correlation:			
White-collar	1.0000	...	...
Blue-collar	.1226 (.0430)	1.0000	...
Military	.0182 (.0997)	.4727 (.0848)	1.0000
	2. Nonpecuniary Values		
Constant	−2,543 (272)	−3,157 (253)	−.0900 (.0448)
Age	...	...	−.0313 (.0057)
	3. Entry Costs		
If positive own experience but not in occupation in previous period	1,182 (285)	1,647 (199)	...
Additional entry cost if no own experience	2,759 (764)	494 (698)	560 (509)
	4. Exit Costs		
One-year military experience	...	...	1,525 (151)

Note.—Standard errors are in parentheses.
* Age is defined as age minus 16.

TABLE 8

ESTIMATED SCHOOL AND HOME PARAMETERS

	School		Home	
Constants:				
Type 1	11,031	(626)	20,242	(608)
Deviation of type 2 from type 1	−5,364	(1,182)	−2,135	(753)
Deviation of type 3 from type 1	−8,900	(957)	−14,678	(679)
Deviation of type 4 from type 1	−1,469	(1,011)	−2,912	(768)
Has high school diploma	804	(137)	...	
Has college diploma	2,005	(225)	...	
Net tuition costs: college	4,168	(838)	...	
Additional net tuition costs: graduate school	7,030	(1,446)	...	
Cost to reenter high school	23,283	(1,359)	...	
Cost to reenter college	10,700	(926)	...	
Age*	−1,502	(111)	...	
Aged 16–17	3,632	(1,103)	...	
Aged 18–20	...		−1,027	(538)
Aged 21 and over	...		−1,807	(568)
Error standard deviation	12,821	(735)	9,350	(576)
Discount factor	.9363 (.0014)			

NOTE.—Standard errors are in parentheses.
* Age is defined as age minus 16.

TABLE 9

ESTIMATED TYPE PROPORTIONS BY INITIAL SCHOOLING LEVEL AND TYPE-SPECIFIC ENDOWMENT RANKINGS

	Type 1	Type 2	Type 3	Type 4
Initial schooling:				
Nine years or less	.0491 (· · ·)	.1987 (.0294)	.4066 (.0357)	.3456 (.0359)
10 years or more	.2343 (· · ·)	.2335 (.0208)	.3734 (.0229)	.1588 (.0183)
Rank ordering:				
School attainment at age 16	1	2	3	4
White-collar skill endowment	1	2	4	3
Blue-collar skill endowment	2	1	4	3
Consumption value of school net of effort cost	1	3	4	2
Value of home production	1	2	4	3

NOTE.—Standard errors are in parentheses.

TABLE 10

MODEL PREDICTIONS VS. CPS CHOICE FREQUENCIES

Age Range	NLSY*	CPS (Year)†	DP-Basic*	DP-Extended†	Approximation*
			White-Collar		
16–19	.043	.064 (1981)	.052	.043	.041
20–23	.190	.187 (1985)	.176	.187	.180
24–26	.344	.345 (1989)	.307	.335	.332
24–27	...	.348 (1989)	.323	.343	.349
28–31	...	.384 (1993)	.365	.375	.443
30–33	...	.413 (1995)	.370	.388	.472
35–44	...	.449 (1995)	.405	.430	.547
			Blue-Collar		
16–19	.171	.265 (1981)	.199	.182	.176
20–23	.430	.432 (1985)	.416	.418	.434
24–26	.475	.472 (1989)	.544	.490	.498
24–27	...	.476 (1989)	.565	.494	.498
28–31	...	.465 (1993)	.616	.539	.495
30–33	...	.460 (1995)	.624	.547	.487
35–44	...	.423 (1995)	.595	.541	.440

* Military is excluded to facilitate comparison with CPS (which is a civilian sample).

† Choice frequencies pertain to whites in the March CPS from the years indicated. We classify a person as working if, over the previous calendar year, he worked at least 35 weeks and, in those weeks, he worked at least 20 hours per week on average. The occupation is that held longest in the previous year.

TABLE 11

SELECTED CHARACTERISTICS AT AGE 24 BY TYPE: NINE OR 10 YEARS INITIAL SCHOOLING

	Initial Schooling 9 Years or Less				Initial Schooling 10 Years or More			
	Type 1	Type 2	Type 3	Type 4	Type 1	Type 2	Type 3	Type 4
Schooling	15.6	10.6	10.9	11.0	16.4	12.5	12.4	13.0
Experience:								
White-collar	.528	.704	.742	.279	1.07	1.06	1.05	.436
Blue-collar	.189	4.05	2.85	1.61	.176	3.65	2.62	1.77
Military	.000	.000	1.35	.038	.000	.000	1.10	.034
Proportion who chose:								
White-collar	.509	.123	.176	.060	.673	.236	.284	.155
Blue-collar	.076	.775	.574	.388	.039	.687	.516	.441
Military	.000	.000	.151	.010	.000	.000	.116	.005
School	.416	.008	.013	.038	.239	.024	.025	.074
Home	.000	.095	.086	.505	.050	.053	.059	.325

NOTE.—Based on a simulation of 5,000 persons.

TABLE 12

EXPECTED PRESENT VALUE OF LIFETIME UTILITY FOR ALTERNATIVE CHOICES AT AGE 16 AND AT AGE 26 BY TYPE ($)

	All Types	Type 1	Type 2	Type 3	Type 4
	Initial Schooling 10 Years or More				
School:					
Age 16	321,008	415,435	394,712	228,350	289,683
Age 26	384,352	499,162	494,107	272,985	314,708
Home:					
Age 16	298,684	380,660	376,945	207,768	274,901
Age 26	426,837	611,167	516,547	291,932	338,653
White-collar:					
Age 16	293,683	372,544	372,733	207,586	262,370
Age 26	439,970	637,616	528,107	303,228	338,967
Blue-collar:					
Age 16	296,736	373,156	377,618	210,699	266,206
Age 26	438,240	617,873	534,578	305,641	342,195
Military:					
Age 16	285,686	350,655	356,202	210,461	261,944
Age 26	415,374	581,996	492,531	298,431	329,938
Maximum over choices:					
Age 16	321,921	415,503	396,108	229,265	291,122
Age 26	445,488	638,820	537,226	308,259	346,695
	Initial Schooling Nine Years or Less				
School:					
Age 16	273,186	387,384	371,369	211,942	276,040
Age 26	308,808	564,590	446,163	243,734	274,979
Home:					
Age 16	260,668	352,274	360,495	197,288	268,047
Age 26	334,643	578,637	468,465	268,815	305,262
White-collar:					
Age 16	253,764	342,833	354,261	196,294	253,686
Age 26	339,093	602,915	474,796	277,488	300,917
Blue-collar:					
Age 16	257,720	343,873	359,370	199,945	257,697
Age 26	344,179	583,895	486,456	282,223	305,520
Military:					
Age 16	251,710	322,293	340,126	199,737	254,386
Age 26	328,916	550,521	447,443	275,660	295,996
Maximum over choices:					
Age 16	275,634	387,384	374,154	213,823	286,311
Age 26	347,741	604,549	487,466	284,073	310,598

NOTE.—Based on a simulation of 5,000 persons.

TABLE 13

RELATIONSHIP OF INITIAL SCHOOLING AND TYPE TO SELECTED FAMILY BACKGROUND CHARACTERISTICS

	Initial Schooling Nine Years or Less and Person Is of Type				Initial Schooling 10 Years or More and Person Is of Type					Expected Present Value of Lifetime Utility at Age 16
	1 (1)	2 (2)	3 (3)	4 (4)	1 (5)	2 (6)	3 (7)	4 (8)	Observations (9)	(10)
All	.010	.051	.103	.090	.157	.177	.289	.123	1,373	307,673
Mother's schooling:										
Non–high school graduate	.004	.099	.177	.161	.038	.141	.276	.103	333	286,642
High school graduate	.011	.043	.086	.071	.143	.210	.305	.131	685	309,275
Some college	.023	.021	.043	.058	.294	.166	.263	.133	152	328,856
College graduate	.007	.005	.049	.023	.388	.151	.222	.154	142	339,593
Household structure at age 14:										
Live with mother only	.001	.062	.133	.119	.123	.137	.297	.128	178	296,019
Live with father only	.026	.037	.088	.120	.062	.180	.378	.106	44	291,746
Live with both parents	.011	.049	.097	.082	.169	.184	.284	.124	1,123	310,573
Live with neither parent	.0001	.090	.154	.184	.037	.175	.275	.085	28	290,469
Number of siblings:										
0	.002	.041	.086	.092	.142	.227	.285	.126	50	310,833
1	.002	.029	.064	.051	.236	.199	.287	.133	261	320,697
2	.016	.048	.104	.063	.191	.157	.275	.146	364	311,053
3	.013	.056	.119	.090	.147	.182	.288	.104	320	306,395
4+	.009	.067	.117	.141	.081	.171	.303	.111	378	296,089
Parental income in 1978:										
$Y \leq \frac{1}{2}$ median*	.002	.078	.155	.181	.071	.132	.221	.161	214	292,565
$\frac{1}{2}$ median $< Y \leq$ median	.007	.053	.120	.103	.103	.173	.328	.113	382	296,372
Median $\leq Y \leq 2 \cdot$ median	.015	.044	.071	.051	.177	.204	.304	.134	446	314,748
$Y \geq 2 \cdot$ median	.014	.025	.024	.021	.479	.167	.182	.087	83	358,404

* Median income in the sample is $20,000.

TABLE 14

EFFECT OF A $2,000 COLLEGE TUITION SUBSIDY ON SELECTED CHARACTERISTICS BY TYPE

	All Types	Type 1	Type 2	Type 3	Type 4
Percentage high school graduates:					
No subsidy	74.8	100.0	68.6	70.2	67.0
Subsidy	78.3	100.0	73.2	74.0	72.2
Percentage college graduates:					
No subsidy	28.3	98.7	11.1	8.6	19.5
Subsidy	36.7	99.5	21.0	17.1	32.9
Mean schooling:					
No subsidy	13.0	17.0	12.1	12.0	12.4
Subsidy	13.5	17.0	12.7	12.5	13.0
Mean years in college:					
No subsidy	1.34	3.97	.69	.59	1.05
Subsidy	1.71	3.99	1.14	1.00	1.58

NOTE.—Subsidy of $2,000 each year of attendance. Based on a simulation of 5,000 persons.

TABLE 15