Dynamic Discrete Choice Models

Zi Wang
HKBU

Spring 2026

Dynamic Discrete Choice

- Last time: demand estimation

$$\sigma_{jt} = \int \frac{\exp\left\{\delta_{jt}\left(\bar{\beta}, \xi_{jt}\right) + x_{jt}'\eta_i\right\}}{1 + \sum_{k=1}^{J} \exp\left\{\delta_{kt}\left(\bar{\beta}, \xi_{kt}\right) + x_{kt}'\eta_i\right\}} dF_{\eta}\left(\eta_i | \Sigma\right) \tag{1}$$

- Discrete choice:
  - Latent variable (value)
  - Idiosyncratic shock
- Dynamic discrete choice: discrete choice $\Rightarrow$ states $\Rightarrow$ future value
  - Investment: Rust (1987); Chen et al. (2023)
  - Occupation and schooling: Keane and Wolpin (1997)
  - Export: Alessandria et al. (2025)
  - Migration: Caliendo, L., Dvorkin, M. and Parro, F. (2019)

Rust Model

- Discrete time t; agent i; finite or infinite horizon
- s_t: vector of state variables; $a_t \in A = \{0, 1, \ldots, J\}$: discrete action
- Markov transition density $G\left(s_{t+1} | s_t, a_t\right)$; usually assume rational expectation
- agent's dynamic programming

$$V\left(s_t\right) = \max_{a \in A} \left\{ u\left(s_t, a\right) + \beta \int V\left(s_{t+1}\right) dG\left(s_{t+1} | s_t, a\right) \right\} \tag{2}$$

- additive separable:

$$u\left(s_t, a\right) = u\left(x_t, a_t\right) + \epsilon(a_t) \tag{3}$$

Rust (1987): Parameterization

- x_t: mileage of an engine; a_t: whether to replace the engine

$$x_{t+1} = \begin{cases} \sim G\left(x'|x_t\right) & \text{if } a_t = 0 \\ = 0 & \text{if } a_t = 1 \end{cases} \tag{4}$$

- $a_t \in \{0, 1\}$; the choice-specific value function

$$v\left(x_t, a_t, \epsilon_t\right) = \begin{cases} u\left(x, 1, \epsilon; \theta\right) + \beta V\left(0\right) & \text{if } a_t = 1 \\ u\left(x, 0, \epsilon, \theta\right) + \beta E_{x', \epsilon' | x_t, \epsilon_t, a_t} V\left(x', \epsilon'\right) & \text{if } a_t = 0 \end{cases} \tag{5}$$

- Parametric assumptions on utility flow:

$$u\left(a, x, \epsilon; \theta\right) = -c\left(\left(1 - a\right) x; \theta\right) - a \times RC + \epsilon(a) \tag{6}$$

where

- $c\left(.\right)$ is the maintenance cost function, increasing in x
- RC: the lumpy fixed cost of adjustment
- $\epsilon(a)$, $a = 0, 1$: structural errors, factors affect replacement choice a_t but unobserved by the econometrician

Rust (1987): Parameterization

- Parameters to be estimated:
  - parameters of maintenance cost function $c(.)$
  - replacement cost RC
  - parameters of mileage transition function $F(x'|x, a)$
- β is not identified; any time series data on (a_t, x_t) could be equally well explained by a myopic model:

$$a_t = \arg\max_{a \in \{0,1\}} \{u(x_t, 0, \epsilon_t), u(x_t, 1, \epsilon_t)\} \tag{7}$$

or a forward-looking model

$$a_t = \arg\max_{a \in \{0,1\}} \{v(x_t, 0, \epsilon_t), v(x_t, 1, \epsilon_t)\} \tag{8}$$

Rust (1987): Econometric Model

- Data: observe $\{a_t, x_t\}$, $t = 1, \ldots, T$ for 62 buses

- Conditional independence implies that:

$$G\left(x', \epsilon' | x, \epsilon, a\right) = G\left(\epsilon' | x'\right) \times G\left(x' | x, a\right) \tag{9}$$

- Likelihood function for a single bus:

$$\begin{aligned} &\ell\left(x_1, \ldots, x_T; a_1, \ldots, a_T | a_0, x_0; \theta\right) \\ &= \prod_{t=1}^{T} Pr\left(a_t, x_t | a_0, x_i, \ldots, a_{t-1}, x_{t-1}; \theta\right) \\ &= \prod_{t=1}^{T} Pr\left(a_t, x_t | a_{t-1}, x_{t-1}; \theta\right) \\ &= \prod_{t=1}^{T} Pr\left(a_t | x_t; \theta\right) \times Pr\left(x_t | x_{t-1}, a_{t-1}; \theta_3\right) \end{aligned} \tag{10}$$

Rust (1987): Estimation

- **Step 1:** estimate θ_3, the parameters of the Markov transition probabilities for mileage, conditional on non-replacement (i.e. $a_t = 0$)

- Example: Discretize: $\Delta x_t \equiv x_{t+1} - x_t$, with

$$\Delta x_t = \begin{cases} [0, 5000) & \text{w/prob } p \\ [5000, 10000) & \text{w/prob } q \\ [10000, \infty) & \text{w/prob } 1 - p - q \end{cases} \tag{11}$$

- $\theta_3 = \{p, q\}$

Rust (1987): Estimation

- **Step 2:** estimate θ, parameters of maintenance cost $c(.)$ and replacement costs
  - **Outer loop:** search over different values of $\hat{\theta}$
  - **Inner loop:** compute the likelihood $Pr\left[a_t | x_t; \hat{\theta}\right]$
- ϵ's are IID TIEV: $\epsilon \sim G_\epsilon$
  - $V(x, \epsilon; \theta) = \max\{v(x, 0, \epsilon; \theta), v(x, 1, \epsilon; \theta)\} = \max\left\{u(x, 1, \epsilon; \theta) + \beta V(0), u(x, 0, \epsilon, \theta) + \beta E_{x', \epsilon' | x_t, \epsilon_t, a_t} V(x', \epsilon')\right\}$
  - Denote $\bar{V}(x; \theta) \equiv \int V(x, \epsilon) \, dG_\epsilon$. Then

$$\bar{V}(x; \theta) = \log\left[\sum_a \exp\left\{u(x, a) + \beta \int \bar{V}\left(x'; \theta\right) dG\left(x' | x, a; \theta_3\right)\right\}\right] \tag{12}$$

  - Then the likelihood can be computed by

$$Pr\left[a_t | x_t; \theta\right] = \frac{\exp\left[u\left(x_t, a_t; \theta\right) + \beta \int \bar{V}\left(x_{t+1}; \theta\right) dG\left(x_{t+1} | x_t, a_t; \theta_3\right)\right]}{\sum_a \exp\left[u\left(x_t, a; \theta\right) + \beta \int \bar{V}\left(x_{t+1}; \theta\right) dG\left(x_{t+1} | x_t, a; \theta_3\right)\right]} \tag{13}$$

Rust Model

- Individual-level:
  - Discrete choices $\Leftarrow$ latent variables
  - Markov transitions in endogenous and exogenous state variables
- "Aggregation":
  - IID TIEV shocks to "smooth" the discrete choices into aggregate shares/probabilities/likelihood: Value function iteration in terms of expected value function
  - Enable the model to capture individual variations
- Estimation:
  - Observed state transition $\Rightarrow$ Parameters for exogenous state transition
  - Likelihood of certain policy (choice) + Observed policy under certain states $\Rightarrow$ Parameters for utility/return

Keane and Wolpin (1997)

- Dynamic model of schooling, work, and occupational choice decisions
  - Human capital investment theory: a potential vehicle for explaining observed patterns of school attendance, work, occupational choice, and wages
  - Data: 11 years of observations on a sample of young men from the 1979 youth cohort of the National Longitudinal Surveys of Labor Market Experience (NLSY)
- Structural approach
  - Estimate parameters that may be of interest in their own right: *e.g.* isolates the quantitative importance of school attainment and occupation-specific work experience in the production of occupation-specific skills
  - Quantify the effect on decisions of altering specific parameters of the environment: *e.g.* the impact of an intervention, such as a college tuition subsidy, on subsequent occupational choice decisions

Keane and Wolpin (1997): Discrete Choices

- Each individual has a finite decision horizon beginning at age 16 and ending at age A

- At age a, an individual chooses among five mutually exclusive and exhaustive alternatives: (i) work in a blue-collar occupation; (ii) work in a white-collar occupation; (iii) work in the military; (iv) attend school; (v) engage in home production

- Choice dummy: $d_m(a) = 1$ for $m = 1, \ldots, 5$; reward per period

$$R(a) = \sum_{m=1}^{5} R_m(a) d_m(a). \tag{14}$$

Keane and Wolpin (1997): Working Alternatives

- Working Alternatives: Reward

$$R_m(a) = \underbrace{w_m(a)}_{\text{wage}} = \underbrace{r_m}_{\text{price of human capital}} \times \underbrace{e_m(a)}_{\text{quantity of human capital}} \tag{15}$$

- Human capital accumulation:

$$e_m(a) = \exp\left[\underbrace{e_m(16)}_{\text{endowment}} + e_{m1} \underbrace{g(a)}_{\text{year of schooling}} + e_{m2} \underbrace{x_m(a)}_{\text{work experience}} + e_{m3} x_m^2(a) + \underbrace{\epsilon_m(a)}_{\text{shock}}\right] \tag{16}$$

Keane and Wolpin (1997): Non-Working Alternatives

- School attendance

$$R_4(a) = e_4(16) - \underbrace{tc_1 \times I\left[g(a) \geq 12\right]}_{\text{cost of college}} - \underbrace{tc_2 \times I\left[g(a) \geq 16\right]}_{\text{cost of graduate school}} + \epsilon_4(a). \quad (17)$$

- Home production

$$R_5(a) = e_5(16) + \epsilon_5(a). \quad (18)$$

Keane and Wolpin (1997): Dynamic Problem

- Shock vector $\epsilon(a) \equiv (\epsilon_1(a), \ldots, \epsilon_5(a))$: joint normal $N(0, \Omega)$; serial uncorrelated
- Initial condition: $g(16)$ and $x_m(16) = 0$ for $m = 1, 2, 3$
- Endowments: $\mathbf{e}(16) = (e_1(16), \ldots e_5(16))$; work experience: $\mathbf{x}(a) = (x_1(a), x_2(a), x_3(a))$
- State vector: $\mathbf{S}(a) = (\mathbf{e}(16), g(a), \mathbf{x}(a), \epsilon(a))$; dynamic problem

$$V(\mathbf{S}(a), a) = \max_{(d_m(\tau))_{\tau=a}^{A}} E\left[\sum_{\tau=a}^{A} \delta^{\tau-a} \sum_{m=1}^{5} R_m(\tau) d_m(\tau) | \mathbf{S}(a)\right]. \tag{19}$$

Keane and Wolpin (1997): Dynamic Programming

- Bellman equation:

$$V\left(\mathbf{S}(a), a\right) = \max_{m=1,\dots,5} \left\{V_m\left(\mathbf{S}(a), a\right)\right\}, \tag{20}$$

  where

$$V_m\left(\mathbf{S}(a), a\right) = \begin{cases} R_m\left(\mathbf{S}(a), a\right) + \delta E\left[V\left(\mathbf{S}(a+1), a+1\right) \middle| \mathbf{S}(a), d_m(a) = 1\right] & \text{if } a < A \\ R_m\left(\mathbf{S}(a), a\right) & \text{if } a = A \end{cases} \tag{21}$$

  and

$$\begin{aligned} x_m(a+1) &= x_m(a) + d_m(a), \quad m = 1, 2, 3 \\ g(a+1) &= g(a) + d_4(a), \quad g(a) \leq \bar{G} \end{aligned} \tag{22}$$

- Joint normal shocks + no additive separability $\Rightarrow$ no closed-form solution $\Rightarrow$ simulate the joint distribution of states and choices

Keane and Wolpin (1997): Likelihood

- For each individual $n = 1, \dots, N$, we observe
  - choices+rewards: $\{d_{nm}(a), w_{nm}(a)d_{nm}(a)\}_{m=1}^{3}$
  - choices: $(d_{nm}(a))_{m=4}^{5}$
- $c(a)$: choice-reward combination at age a; $\bar{\mathbf{S}}(a) = \{\mathbf{e}(16), g(a), \mathbf{x}(a)\}$
- Serial independence implies that

$$Pr\left[c(16), \dots, c\left(\bar{a}\right) | g(16), \mathbf{e}(16)\right] = \prod_{a=16}^{\bar{a}} Pr\left[c(a) | \bar{\mathbf{S}}(a)\right]. \quad (23)$$

- Two generalizations:
  - Heterogeneous initial endowments: $\mathbf{e}_k(16) = \{e_{mk}(16) : m = 1, \dots, 5\}$ where $k = 1, \dots, K$ with prob. π_k
  - Heterogeneous initial schooling: $g_n(16)$ is exogenous conditional on initial endowments: prob. $\pi_{k|g_n(16)}$

Keane and Wolpin (1997): Data Patterns

TABLE 1

CHOICE DISTRIBUTION: WHITE MALES AGED 16–26

Age	School	Home	White-Collar	Blue-Collar	Military	Total
	Choice					
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.

Keane and Wolpin (1997): Data Patterns

TABLE 2

TRANSITION MATRIX: WHITE MALES AGED 16–26

CHOICE ($t-1$)	CHOICE (t) 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

Keane and Wolpin (1997): Data Patterns

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			

Keane and Wolpin (1997): Data Patterns

TABLE 4

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

Age	Mean Wage: 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.

Keane and Wolpin (1997): Estimates

TABLE B1

ESTIMATES OF THE BASIC MODEL

A. OCCUPATION-SPECIFIC PARAMETERS

	White-Collar		Blue-Collar		Military	
Skill functions:						
Schooling	.0938	(.0014)	.0189	(.0014)	.0443	(.0027)
White-collar experience	.1170	(.0015)	.0674	(.0017)	...	
Blue-collar experience	.0748	(.0017)	.1424	(.0011)	...	
Military experience	.0077	(.0007)	.1021	(.0021)	.3391	(.0122)
"Own" experience squared/100	−.0461	(.0032)	−.1774	(.0041)	−2.9900	(.2156)
Constants:						
Type 1	8.8043	(.0124)	8.9156	(.0126)	8.4704	(.0234)
Deviation of type 2 from type 1	−.0668	(.0047)	.2996	(.0094)	...	
Deviation of type 3 from type 1	−.4221	(.0100)	−.1223	(.0079)	...	
Deviation of type 4 from type 1	−.4998	(.0176)	.0756	(.0058)	...	
True error standard deviation	.3301	(.0077)	.3329	(.0070)	.3308	(.0156)
Measurement error standard deviation	.4133	(.0065)	.3089	(.0055)	.1259	(.0166)
Error correlation matrix:						
White-collar	1.0010	(· · ·)				
Blue-collar	−.3806	(.0252)	1.0000	(· · ·)		
Military	−.3688	(.0245)	.4120	(.0505)	1.0000	(· · ·)

Keane and Wolpin (1997): Estimates

B. School and Home Parameters

	School		Home
Constants:			
Type 1	43,948 (850)		16,887 (413)
Deviation of type 2 from type 1	−26,352 (757)		215 (377)
Deviation of type 3 from type 1	−30,541 (754)		−16,966 (542)
Deviation of type 4 from type 1	226 (594)		−13,128 (1,000)
Net tuition costs:			
College	2,983 (156)		. . .
Graduate school	26,357 (737)		. . .
Error standard deviation	2,312 (105)		13,394 (460)
Discount factor		.7870 (.0048)	

C. Type Proportions by Initial School Level and Type-Specific Endowment Rankings

	Type 1	Type 2	Type 3	Type 4
Initial schooling:				
Nine years or less	.1751 (· · ·)	.2396 (.0172)	.5015 (.0199)	.0838 (.0125)
10 years or more	.0386 (· · ·)	.4409 (.0344)	.4876 (.0350)	.0329 (.0131)
Rank ordering:				
White-collar	1	2	3	4
Blue-collar	3	1	4	2
Schooling	2	3	4	1
Home	2	1	4	3

Note.—Standard errors are in parentheses.

Keane and Wolpin (1997): Within-Sample Model Fit

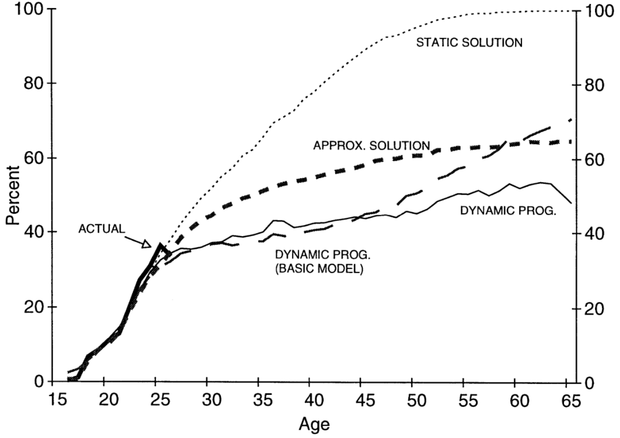

FIG. 1.—Percentage white-collar by age

Keane and Wolpin (1997): Within-Sample Model Fit

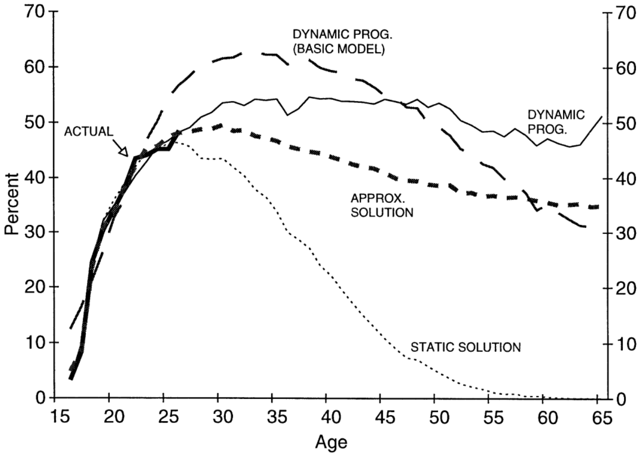

FIG. 2.—Percentage blue-collar by age

Keane and Wolpin (1997): Within-Sample Model Fit

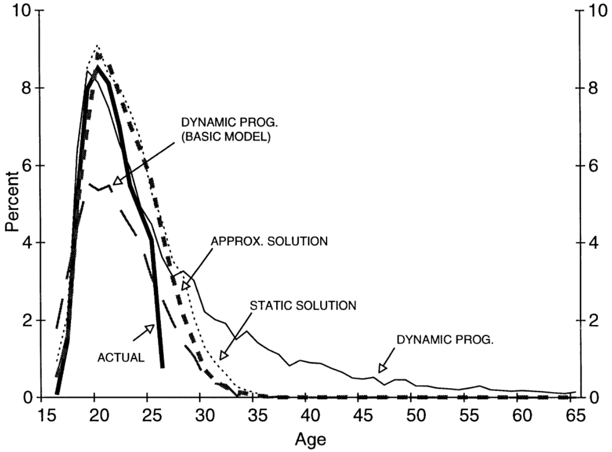

FIG. 3.—Percentage in the military by age

Keane and Wolpin (1997): Within-Sample Model Fit

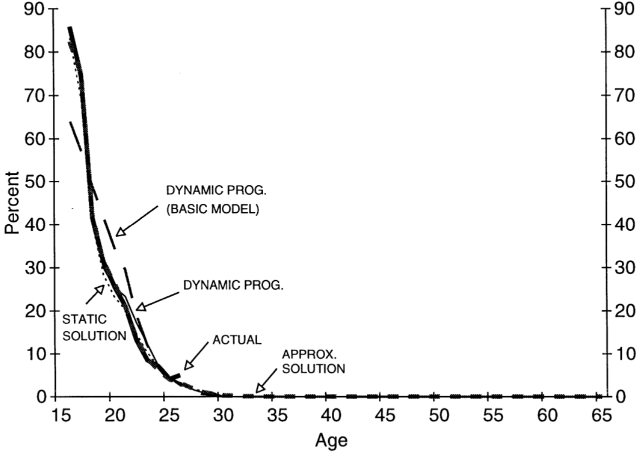

FIG. 4.—Percentage in school by age

Keane and Wolpin (1997): Within-Sample Model Fit

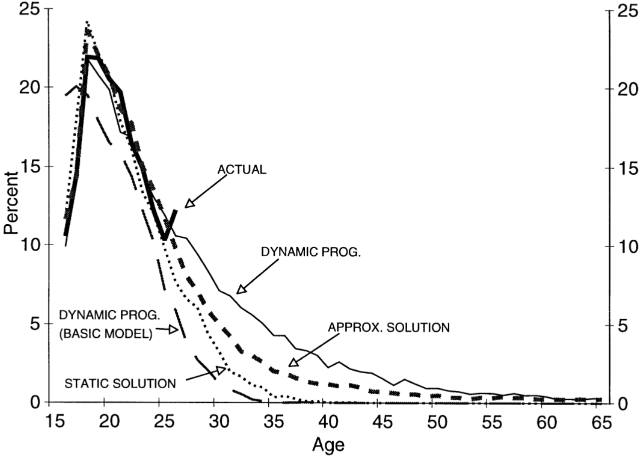

FIG. 5.—Percentage at home by age

Keane and Wolpin (1997): Counterfactuals

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.

Keane and Wolpin (1997): Counterfactuals

TABLE 15

DISTRIBUTIONAL EFFECTS OF A $2,000 COLLEGE TUITION SUBSIDY

	Type 1	Type 2	Type 3	Type 4
Mean expected present value of lifetime utility at age 16:				
No subsidy	413,911	391,162	225,026	286,311
Subsidy	419,628	392,372	226,313	288,109
Gross gain	5,717	1,210	1,287	1,798
Net gain:				
Subsidy to all types*	3,513	−994	−917	−406
Subsidy to types 2, 3, and 4†	−1,134	76	153	664
Subsidy to types 3 and 4‡	−862	−862	425	936

* The per capita cost of the subsidy program is $2,204.
† The per capita cost of the subsidy program is $1,134.
‡ The per capita cost of the subsidy program is $862.

Summary

- Explore the data through the lens of model
  - Identification: Data variations $\Rightarrow$ Model's parameters

- Model fit:
  - Within sample: combinations of states and choices of young men

  - Out-of-sample: combinations of states and choices of "old" men (not used in estimation)

  - How to evaluate model fit? Predictive power?

- Counterfactuals: partial equilibrium
  - Policies $\Rightarrow$ Aggregate variables $\Rightarrow$ Latent variables $\Rightarrow$ Individual choices

  - Quantify the importance of different channels: shut down one channel and re-do counterfactuals