

Decision under uncertainty and dynamic discrete choices: application to single-car owners in France

Matthieu de Lapparent ¹ Giulia Cernicchiaro ²

¹Université Paris-Est, IFSTTAR

²SMART

20 June 2014

Purpose: how long to use a car and how much to it?

- Develop an optimal “use and stop” problem for single-car owners
- Household i determines simultaneously the optimal ownership horizon $\bar{t}_i \in \{1, \dots, T\}$ of its car and the sequence of mileages $\mathbf{m}_{i,\cdot} = (m_{i,1}, \dots, m_{i,T})$ to be driven
- All decisions, car disposal or car keeping (d) and use (m), are made as a function of the state of the household and its environment at the beginning of each period. The vector of state variables is labeled $\mathbf{z}_{i,t}$.
- Application of conditional Logit dynamic discrete choice (Rust, 1985)

Framework I

Households' preferences over possible sequences of state variables can be represented by a time separable discounted (indirect) utility function

$$\sum_{t=1}^T \gamma^t u(\mathbf{z}_{i,t}, d(\mathbf{z}_{i,t}), m(\mathbf{z}_{i,t}))$$

where

- γ : discounting factor
- $u(\mathbf{z}_{i,t}, d(\mathbf{z}_{i,t}), m(\mathbf{z}_{i,t}))$: per-period utility function.

Framework II

At the beginning of each period t :

- The information available for the household is the outcomes of state variables for the period;
- It chooses whether to dispose ($d_{i,t} = 1, D$) its car or to keep ($d_{i,t} = 0, K$) it
- If kept, it chooses an amount of mileage $m_{i,t}$. If it disposes it then no mileage is driven ($m_{i,t} = 0$), it receives a sell-off/scrap value and the decision process stops

Framework III

- The decisions at period t affect the evolution of future values of the state variables, but the household faces uncertainty about these future values.
- The beliefs of household i about sequences of states are modeled by a Markov transition distribution function $G_z(\mathbf{z}_{i,t+1}|\mathbf{z}_{i,t}, d_{i,t})$.
- For simplicity, we note $d_{i,t} = d(\mathbf{z}_{i,t})$ and $m_{i,t} = m(\mathbf{z}_{i,t})$.

Framework IV

The sequence of discrete decisions is made to maximize the expected utility with respect to the distribution of the sequences of state variables:

$$\max_{\bar{t}_i} \left\{ \mathbb{E}_{\mathbf{z}_{i,\cdot}} \left(\sum_{t=1}^{T_i} \gamma^t u(\mathbf{z}_{i,t}, d_{i,t}, m_{i,t}) \mid \mathbf{z}_{i,0} \right) \right\} \quad (1)$$

where $\mathbf{z}_{i,0}$ is a set of initial conditions.

General modeling assumptions I

- The optimal demands for mileages $m^*(\mathbf{z}_{i,t})$ and the decision to keep or dispose the car, $d^*(\mathbf{z}_{i,t})$, are the arguments that maximize utility.
- The set of state variables can be separated in two subsets, $\mathbf{z}_{i,t} = \{\mathbf{x}_{i,t}, \boldsymbol{\epsilon}_{i,t}\}$.
- $\mathbf{x}_{i,t}$ is the set of observed state variables. Here, $\mathbf{x}_{i,t} = \{a_{i,t}, k_{i,t}, y_{i,t}, p_{i,t}\}$. All these variable are considered to take discrete values

General modeling assumptions II

- $\epsilon_{D,i,t}, \epsilon_{K,i,t}$ are two unobserved state variables that are assumed to be iid with cdf $\exp(-\exp(-\epsilon_{d,i,t}))$.
- It is also assumed that the unobserved state variables are modelled as additive shocks
- Conditional independence:
$$G_z(\mathbf{z}_{i,t+1} | \mathbf{z}_{i,t}, d_{i,t}) = G_x(\mathbf{x}_{i,t+1} | \mathbf{x}_{i,t}, d_{i,t}) H_\epsilon(\boldsymbol{\epsilon}_{i,t+1} | d_{i,t})$$
- G_z with a discrete support (computational “easiness”)

Bellman representation

- The intertemporal optimization problem can be formulated as a sequential decision problem:

$$V(\mathbf{z}_{i,t}) = \max \left\{ v_D(\mathbf{x}_{i,t}) + \epsilon_{D,i,t}, v_K(\mathbf{x}_{i,t}) + \gamma \mathbb{E}_{\mathbf{z}_{i,t+1}} [V(\mathbf{z}_{i,t+1}) | \mathbf{z}_{i,t}] + \epsilon_{K,i,t} \right\}.$$

Utility functions

- The observed state variables affect differently the utility functions related to car disposal and car keeping:

$$v_D(\mathbf{x}_{i,t}) = \alpha \frac{P_i}{1 + k_{i,t} a_{i,t}},$$

and

$$v_K(\mathbf{x}_{i,t}) = \beta_1 + \sum_{q=2}^Q \beta_q \mathbb{I}(y_{i,t} = q) + \beta_p c_i p_{i,t} + \beta_k k_{i,t}.$$

Observed state variables: age and cumulated mileage

- Age and cumulated mileage at the beginning of a period t are defined as deterministic state variables:

$$k_{i,t} = k_{i,t-1} + m_{i,t-1}.$$

and

$$a_{i,t} = a_{i,t-1} + 1.$$

- Cumulated mileage is a deterministic function of the chosen mileage at the beginning of period $t - 1$, which was optimally chosen as a function of the other state variables.

Observed state variables: mileage, I

- Mileage to be driven at date t when not disposing its car is modelled as a function of income, fuel price, and cumulated mileage at the beginning of the period:

$$\ln \left(1 + m_{i,t}^* \right) = \theta_0 + \sum_{j=2}^Q \theta_j \mathbb{I}(y_{i,t} = j) + \theta_p \ln(c_i p_{i,t}) + \theta_k \ln(1 + k_{i,t}) + \epsilon_{m,i,t}.$$

- choice of a level of mileage at the beginning of period t affects the expected maximum value of the utility function of period $t + 1$ through the cumulated mileage state variable
- $\epsilon_{m,i,t}$ is assumed to be identically and independently distributed with a Logistic distribution which standard deviation is ω_m .

Observed state variables: mileage, II

- Observed mileage: a discrete representation of a continuous choice; better depicted by a discrete distribution
- Observed values are grouped into predetermined intervals that form together a discrete set of choices and for which “representative” values are defined: $m_{i,t} \in \{m_1, \dots, m_B\}$.

Observed state variables: mileage, III

The probability to sample a household i that reports a mileage m_b at the beginning of period t when its car kept one more period is

$$\Pr(m_{i,t} = m_b | k_{i,t}, y_{i,t}, p_{i,t}, c_i, d_{i,t} = 0; \boldsymbol{\theta}, \omega_m) = \frac{1}{1 + \exp\left(\frac{\bar{m}_b - \theta_0 - \sum_{j=2}^Q \theta_j \mathbb{I}(y_{i,t}=j) - \theta_p \ln(c_i p_{i,t}) - \theta_k \ln(1+k_{i,t})}{\omega_m}\right)} \frac{1}{1 + \exp\left(\frac{\bar{m}_{b-1} - \theta_0 - \sum_{j=2}^Q \theta_j \mathbb{I}(y_{i,t}=j) - \theta_p \ln(c_i p_{i,t}) - \theta_k \ln(1+k_{i,t})}{\omega_m}\right)}$$

Observed state variables: fuel prices, I

- Fuel type of an owned car is given at the initial period: either diesel or petrol.

$$p_{i,t} = \mathbb{I}(\text{fuel type of } i \text{ is petrol}) p_{\text{petrol},t} + \mathbb{I}(\text{fuel type of } i \text{ is diesel}) p_{\text{diesel},t},$$

- Different households having cars with the same fuel type are faced with the same market price

Observed state variables: fuel prices, II

- Expectations about evolution of fuel prices are based on a discrete representation of their respective processes (households are sensitive stage by stage).
- Their dynamics are not stationnary over time.
 $\forall g \in \{\text{petrol, diesel}\},$

$$p_{g,t} = p_{g,t-1} + \epsilon_{g,t}$$

where $\epsilon_{g,t} \stackrel{iid}{\rightarrow} \mathcal{N}(0, \sigma_g^2)$.

Observed state variables: fuel prices, II

- Let $(\bar{p}_{g,0}, \dots, \bar{p}_{g,R})$ be a sequence of R predetermined values over the real line (from 0 to 4 € by step of 1 cent).
- $\tilde{p}_{g,r} = (\bar{p}_{g,r} + \bar{p}_{g,r+1})/2$. Then

$$\Pr(p_{g,t} = \tilde{p}_{g,r} | p_{g,t-1} = \tilde{p}_{g,l}, d_{i,t} = 0; \sigma_g) = \Phi\left(\frac{\bar{p}_{g,r} - \tilde{p}_{g,l}}{\sigma_g}\right) - \Phi\left(\frac{\bar{p}_{g,r-1} - \tilde{p}_{g,l}}{\sigma_g}\right)$$

Observed state variables: income, I

- Income class which is observed as a categorical variable (Q income classes): the actual level of income $y_{i,t}^*$ is a latent variable for which we assume a very simple law of motion:

$$\ln(y_{i,t}^*) = \beta_0 + \sum_{q=2}^Q \beta_q \mathbb{I}(y_{i,t-1} = q) + \epsilon_{y,i,t}$$

where $\epsilon_{y,i,t} \stackrel{iid}{\rightarrow} \mathcal{L}ogistic(0, \tau_y)$.

- At date t , the income class of household i is $y_{i,t} = j$ if and only if $\bar{y}_{j-1} \leq \ln(y_{i,t}^*) < \bar{y}_j$.
- $\bar{y}_0, \dots, \bar{y}_J$ are predetermined point values.

Observed state variables: income, II

- The transition probability of household i to income class j at date t from any class at date $t - 1$ is therefore

$$\Pr(y_{i,t} = j | y_{i,t-1}, d_{i,t} = 0; \beta, \tau_y) = \frac{1}{1 + \exp\left(-\frac{\bar{y}_j - \delta_0 - \sum_{q=2}^Q \delta_q \mathbb{I}(y_{i,t-1}=q)}{\tau_y}\right)} - \frac{1}{1 + \exp\left(-\frac{\bar{y}_{j-1} - \delta_0 - \sum_{q=2}^Q \delta_q \mathbb{I}(y_{i,t-1}=q)}{\tau_y}\right)}$$

where, by convention, $\bar{y}_0 \equiv -\infty$ and $\bar{y}_Q \equiv +\infty$.

Estimation

- Estimation of the structural parameters $\lambda = \{\alpha, \beta, \theta, \delta, \omega_m, \sigma_{\text{petrol}}, \sigma_{\text{diesel}}, \tau_y\}$, and the discount factor γ
- What is observed for each household i are the sequences of choices (keep/dispose and mileages) and states but also initial conditions. Except for fuel prices, data are observed only when car is owned.

Likelihood function

- Panel is not cylindric. The likelihood function of the observed sample may be written as

$$\sum_{i=1}^n \ln \ell(\boldsymbol{\lambda}, \gamma | \mathbf{d}_{i,\cdot}, \mathbf{m}_{i,\cdot}, \mathbf{y}_{i,\cdot}, \mathbf{p}_i)$$

where

$$\begin{aligned} \ell(\boldsymbol{\lambda}, \gamma | \mathbf{d}_{i,\cdot}, \mathbf{m}_{i,\cdot}, \mathbf{y}_{i,\cdot}, \mathbf{p}_i) = & \\ & \Pr(d_{i,\bar{t}_i} = 1 | y_{i,\bar{t}_i}, a_{i,\bar{t}_i}, k_{i,\bar{t}_i}, p_{i,\bar{t}_i}, \mathbf{x}_{i,0}; \boldsymbol{\alpha}, \boldsymbol{\beta}) \\ & \prod_{t=\bar{t}_i}^{\bar{t}_i-1} \Pr(d_{i,t} = 0 | m_{i,t}, y_{i,t}, a_{i,t}, k_{i,t}, p_{i,t}, \mathbf{x}_{i,0}; \boldsymbol{\lambda}, \gamma) \\ & \prod_{t=\bar{t}_i}^{\bar{t}_i-1} \Pr(m_{i,t} | y_{i,t}, k_{i,t}, p_{i,t}, d_{i,t} = 0, \mathbf{x}_{i,0}; \boldsymbol{\theta}, \omega_m) \\ & \prod_{t=\bar{t}_i}^{\bar{t}_i} \Pr(p_{i,t} | p_{i,t-1}, d_{i,t} = 0, \mathbf{x}_{i,0}; \sigma_{\text{petrol}}, \sigma_{\text{diesel}}) \\ & \prod_{t=\bar{t}_i}^{\bar{t}_i} \Pr(y_{i,t} | y_{i,t-1}, d_{i,t} = 0, \mathbf{x}_{i,0}; \boldsymbol{\beta}, \tau_y) \end{aligned}$$

Choice probabilities

One obtains as choice probabilities, assuming that households are utility maximizers,

$$\Pr(d_{i,t} = 1 | y_{i,t}, a_{i,t}, k_{i,t}, p_t, \mathbf{x}_{i,0}; \alpha, \beta) = \frac{\exp(v_D(\mathbf{x}_{i,t}))}{\exp(v_D(\mathbf{x}_{i,t})) + \exp(v_K(\mathbf{x}_{i,t}))},$$

and

$$\Pr(d_{i,t} = 0 | y_{i,t}, a_{i,t}, k_{i,t}, p_t, \mathbf{x}_{i,0}; \lambda, \gamma) = \frac{\exp(v_K(\mathbf{x}_{i,t}) + \gamma \mathbb{E}_{\mathbf{z}_{i,t+1}} [V(\mathbf{z}_{i,t+1}) | \mathbf{z}_{i,t}])}{\exp(v_K(\mathbf{x}_{i,t}) + \gamma \mathbb{E}_{\mathbf{z}_{i,t+1}} [V(\mathbf{z}_{i,t+1}) | \mathbf{z}_{i,t}]) + \exp(v_D(\mathbf{x}_{i,t}))},$$

DP problem I

In order to evaluate the log-likelihood function for particular values of λ and γ , the dynamic programming problem needs to be solved exactly, or its solution approximated in some way.

Under our assumptions, one obtains

$$\mathbb{E}_{\mathbf{z}_{i,t}} [V(\mathbf{z}_{i,t}) | \mathbf{z}_{i,t-1}] = \sum_{q=1}^Q \sum_{r=1}^R \left\{ \begin{array}{l} \Pr(y_{i,t} | y_{i,t-1}, d_{i,t} = 0, \mathbf{x}_{i,0}; \beta, \tau_y) \times \\ \Pr(p_{i,t} | p_{i,t-1}, d_{i,t} = 0, \mathbf{x}_{i,0}; \sigma_{\text{petrol}}, \sigma_{\text{diesel}}) \times \\ \ln(\exp(v_K(\mathbf{x}_{i,t}) + \gamma \mathbb{E}_{\mathbf{z}_{i,t+1}} [V(\mathbf{z}_{i,t+1}) | \mathbf{z}_{i,t}]) \\ + \exp(v_D(\mathbf{x}_{i,t}))) \end{array} \right\}$$

DP problem II

- Backward induction starting from the last date of observation and the terminal condition
$$\mathbb{E}_{\mathbf{z}_{i,t+1}} [V(\mathbf{z}_{i,t+1}) | \mathbf{z}_{i,t}] = 0 \text{ if } d_t = 1$$
- Estimation of $\boldsymbol{\lambda}, \gamma$ is made in two steps (not efficient):
 - estimate the parameters of the transition probability distributions
 - estimate the parameters of the dynamic programming problem given the transition probability distributions:
 - ▶ inner step: evaluating Bellman equation for the current value of $\boldsymbol{\lambda}, \gamma$
 - ▶ outer step: finding a new value of $\boldsymbol{\lambda}, \gamma$ by iterating over the partial log-likelihood function that regards choice probabilities
 - discount factor is fixed

Data

- Data are drawn from the French “Parc Auto” (rotative) panel survey over the period 2000-2007
- Data are completed by drawing time series on fuel prices in France over the time period in the “DIREM” database.
- Population: households who owned at most one car and who disposed it during the period (51.57% of observations).
- 310 observed HH-cars and a total of 1151 observations
- For each HH-car were computed mileages, cumulated mileages, fuel consumption, age, income class, fuel type, purchase price.

Results: fuel prices

We have actually very little information on fuel prices. Despite the very small samples, the baseline assumption was not rejected.

Table: Estimates: fuel prices

Label	Estimate
Variance σ_{petrol}^2	0.026
Variance σ_{diesel}^2	0.037

Results: income

The richer household i was at date $t - 1$, the larger the probability to belong to a higher class of income at date t .

Table: Estimates: income

Label	Estimate	Std. Dev.	T-stat.
Intercept	1.870	0.026	72.350
Previous inc. class is [15.2; 19.1[K€	0.451	0.035	12.810
Previous inc. class is [19.1; 22.9[K€	0.748	0.035	21.580
Previous inc. class is [22.9; 26.7[K€	0.953	0.039	24.250
Previous inc. class is [26.7; 30.5[K€	1.159	0.036	32.610
Previous inc. class is [30.5; 38.1[K€	1.304	0.033	39.130
Previous inc. class is [38.1; 45.7[K€	1.449	0.037	39.510
Previous inc. class is [45.7; 61[K€	1.614	0.042	38.260
Previous inc. is ≥ 61 K€	1.911	0.085	22.510
Variance τ_y^2	0.019	0.001	13.930
Log-lik. at convergence	1152.781		

Results: demand for mileage

Table: Estimates: demand for mileage

Label	Estimate	Std. Dev.	T-stat.
Intercept	3.774	0.304	12.420
Inc. class is [15.2; 19.1[K€	0.069	0.063	1.100
Inc. class is [19.1; 22.9[K€	0.244	0.060	4.040
Inc. class is [22.9; 26.7[K€	0.220	0.074	2.970
Inc. class is [26.7; 30.5[K€	0.334	0.067	5.020
Inc. class is [30.5; 38.1[K€	0.414	0.062	6.710
Inc. class is [38.1; 45.7[K€	0.470	0.071	6.660
Inc. class is [45.7; 61[K€	0.418	0.086	4.850
Inc. is ≥ 61 K€	0.261	0.153	1.700
Average fuel exp. θ_p	-0.593	0.070	-8.460
Cum. mileage θ_k	0.239	0.023	10.550
Variance ω_m^2	0.088	0.005	16.300
Log-lik. at convergence	1780.435		

Results: probability to keep, $\gamma = 0$

Table: Estimates: probability to keep

Label	Estimate	Std. Dev.	T-stat.
Intercept	1.887	0.355	5.320
Inc. class is [15.2; 19.1[K€	-0.101	0.238	-0.420
Inc. class is [19.1; 22.9[K€	-0.136	0.236	-0.580
Inc. class is [22.9; 26.7[K€	-0.319	0.254	-1.260
Inc. class is [26.7; 30.5[K€	-0.108	0.277	-0.390
Inc. class is [30.5; 38.1[K€	-0.452	0.236	-1.910
Inc. class is [38.1; 45.7[K€	-0.303	0.292	-1.040
Inc. class is [45.7; 61[K€	0.101	0.334	0.300
Inc. is ≥ 61 K€	1.346	0.800	1.680
Average fuel exp. β_p	-0.010	0.004	-2.800
Cum. mileage β_k	-0.003	0.001	-2.290
Scrap/sell-off value α	4.649	1.199	3.880
Discount factor γ	0		
Log-lik. at convergence	-636.075		

Results: probability to keep, $\gamma = 0.5$

Table: Estimates: probability to keep

Label	Estimate	Std. Dev.	T-stat.
Intercept	1.422	0.380	3.744
Inc. class is [15.2; 19.1[K€	0.005	0.259	0.019
Inc. class is [19.1; 22.9[K€	-0.059	0.257	-0.231
Inc. class is [22.9; 26.7[K€	-0.168	0.276	-0.610
Inc. class is [26.7; 30.5[K€	0.067	0.299	0.225
Inc. class is [30.5; 38.1[K€	-0.551	0.263	-2.100
Inc. class is [38.1; 45.7[K€	-0.557	0.352	-1.582
Inc. class is [45.7; 61[K€	-0.706	0.484	-1.459
Inc. is ≥ 61 K€	10.734	0.758	14.155
Average fuel exp. β_p	-0.007	0.004	-1.945
Cum. mileage β_k	-0.002	0.001	-1.286
Scrap/sell-off value α	2.847	1.372	2.075
Discount factor γ	0.5		
Log-lik. at convergence	-500.606		

Results: probability to keep, $\gamma = 0.99$

Table: Estimates: probability to keep

Label	Estimate	Std. Dev.	T-stat.
Intercept	0.790	0.426	1.855
Inc. class is [15.2; 19.1[K€	0.081	0.290	0.279
Inc. class is [19.1; 22.9[K€	-0.014	0.287	-0.048
Inc. class is [22.9; 26.7[K€	-0.035	0.306	-0.114
Inc. class is [26.7; 30.5[K€	0.125	0.335	0.374
Inc. class is [30.5; 38.1[K€	-0.665	0.313	-2.122
Inc. class is [38.1; 45.7[K€	-0.920	0.479	-1.920
Inc. class is [45.7; 61[K€	-1.802	1.148	-1.569
Inc. is ≥ 61 K€	14.379	0.769	18.690
Average fuel exp. β_p	-0.004	0.004	-0.991
Cum. mileage β_k	-0.001	0.001	-0.525
Scrap/sell-off value α	1.917	1.534	1.250
Discount factor γ	0.99		
Log-lik. at convergence	-384.519		

Further extensions

- Persistent unobserved heterogeneity (e.g. individual random effects);
- Serial correlation in unobserved state variables;
- Additional unobserved state variables;
- Latent state variables: POMDP;
- Infinite horizon;
- Dynamic games
- ...

References

- Berkovec, J. and S. Stern (1991), Job Exit Behavior of Older Men. *Econometrica*, 59, 189-210
- Rust, J. and C. Phelan (1997), How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets. *Econometrica*, 65, 781-832.
- Karlstrom, A., Palme, M. and I. Svensson (2004), A dynamic programming approach to model the retirement behaviour of blue-collar workers in Sweden. *Journal of Applied Econometrics*, 19, 795-807.
- Keane, M. and K. Wolpin (1997), The career decisions of young men. *Journal of Political Economy*, 105, 473-522.

References

- Eckstein, Z. and K. Wolpin (1999), Why Youth Drop out of High School: The Impact of Preferences, Opportunities and Abilities. *Econometrica*, 67, 1295-1339
- Das, S., (1992), A Micro-econometric Model of Capital Utilization and Retirement: The Case of the Cement Industry. *Review of Economic Studies*, 59, 277-297.
- Rust, J. and G. Rothwell (1995), Optimal Response to a Shift in Regulatory Regime: The Case of the US Nuclear Power Industry. *Journal of Applied Econometrics*, 10, 75-118

References

- Aguirregabiria, V. and P. Mira (2010), Dynamic Discrete Choice Structural Models: A Survey, *Journal of Econometrics*, 156, 38-67
- Rust, J. (1994), Structural estimation of Markov decision processes. In R. E. Engle and McFadden (eds.) *Handbook of Econometrics Volume 4*, North-Holland. Amsterdam.
- Pakes, A. (1994), Dynamic structural models, problems and prospects, in C. Sims (ed.) *Advances in Econometrics. Sixth World Congress*, Cambridge University Press
- Magnac, T. and D. Thesmar (2002), Identifying Dynamic Discrete Decision Processes. *Econometrica*, 70, 801-816
- and so many others...