INTRODUCTION TO DYNAMIC DISCRETE CHOICE MODELS: FROM STATIC TO DYNAMIC

Montreal, June 10, 2015

EMMA FREJINGER

Assistant professor

Holder of the CN Chair on Intermodal Transportation

Member of CIRRELT

Department of Computer Science and Operations Research (DIRO)

Université de Montréal

Université de Montréal

CIRRELT

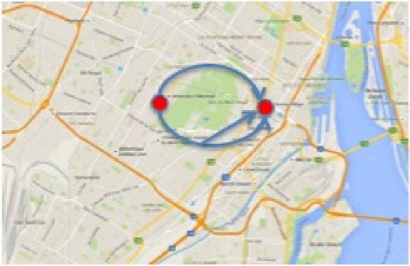
CHAIRE CN INTERMODALITÉ DES TRANSPORTS

OUTLINE

- Introduction to route choice modelling
- Recursive logit
- Rust's model
- Maximum likelihood estimation

INTRODUCTION

- Route choice models play an important role in different transport related applications (e.g. traffic simulation and evaluation of infrastructure investments)



INTRODUCTION

- Choice set of three paths $C_n = \{1,2,3\}$
- Travel time is static and deterministic and is the only attribute observed by the modeller
- $V_1 = V_2 = V_3 = V_j = \beta T_j$
- Static logit model $U_j = V_j + \varepsilon_j$

$$P(1|C_n) = P(2|C_n) = P(3|C_n) = \frac{e^{V_1}}{\sum_{j \in C_n} e^{V_j}}$$

INTRODUCTION

- Transport networks are large and there are many alternatives for each origin-destination pair
- Several challenges associated with route choice modelling: choice sets are unknown and path utilities are correlated
- Addressing the choice set issue: we can reformulate the path choice problem as a sequence of link choices, we call it recursive logit (Fosgerau et al., 2013, Mai et al. 2015)
- Recursive logit for a static and deterministic network is a simple case of a dynamic discrete choice model

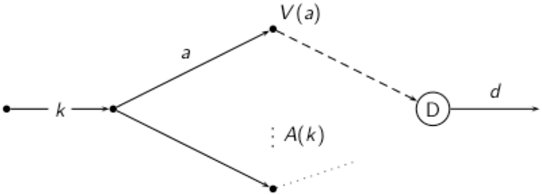
RECURSIVE LOGIT

- Utilities are associated with links $a \in A$ and the deterministic part is $-T_a$, here -4 (upper link) and -2 for the others
- Logit over outgoing links at origin $\frac{e^{-2}}{e^{-2}+e^{-4}} = 0.88$
- Recursive logit $\frac{e^{-2-0.57}}{e^{-2-0.57}+e^{-4}} = 0.50$

RECURSIVE LOGIT

- Directed connected network $G = (A, \mathcal{V})$
- Deterministic attributes
- A state $k \in A$ is a link in the network so that turn attributes can be included
- Choice of one outgoing link (action) $a \in A(k)$ at the sink node of k
- Path: $(k_0, .., k_I)$ with $k_{i+1} \in A(k_i)$ for all $i < I$
- From now on
  - All equations are destination specific
  - Notation for individuals omitted
  - Static network is assumed

RECURSIVE LOGIT



- Infinite horizon problem with an absorbing state d (destination)
- $\widetilde{A} = A \cup d$
- $u(a|k) = v(a|k) + \mu\varepsilon(a)$, $\varepsilon(a)$ i.i.d. EV type 1
- $v(a|k) = v(x_{a|k};\beta) < 0 \ \forall a,k \in A$, $v(d|k) = 0$

RECURSIVE LOGIT

- Traveler chooses next link given current state, stochastic process having Markov property
- **Next state given with certainty by the action**
- Traveler observes $\varepsilon(a)\ \forall a \in A(k)$, chooses action a that maximizes sum of $u(a|k)$ and expected maximum downstream utility $V(a)$

RECURSIVE LOGIT

- Bellman equations

$$V(k) = E\left[\max_{a\in A(k)} \left(v(a|k) + V(a) + \mu\varepsilon(a)\right)\right] \;\forall\; k \in A$$

- Infinite horizon but no discounting!

CHAIRE CN INTERMODALITÉ DES TRANSPORTS

RECURSIVE LOGIT

- Probability of choosing link a given state k

$$P(a|k) = \frac{e^{\frac{1}{\mu}(v(a|k)+V(a))}}{\sum_{a' \in A(k)} e^{\frac{1}{\mu}(v(a'|k)+V(a'))}}$$

RECURSIVE LOGIT

- Expected maximum utility corresponds to the "logsum"

$$V(k) = \begin{cases} \mu \ln \sum_{a \in A} \delta(a|k) e^{\frac{1}{\mu}(v(a|k)+V(a))} & \forall k \in A \\ 0 & k = d \end{cases}$$

$\delta(a|k) = 1$ if $a \in A(k)$ and zero otherwise

- Note that denominator in $P(a|k)$ is $e^{\frac{1}{\mu}V(k)}$

PROPERTIES

- Solving the Bellman equations

$$e^{\frac{1}{\mu}V(k)} = \begin{cases} \sum_{a\in A} \delta(a|k) e^{\frac{1}{\mu}(v(a|k)+V(a))} & \forall\ k \in A, \\ 1 & k = d \end{cases}$$

PROPERTIES

- System of linear equations

$$\mathbf{z} = \mathbf{M}\mathbf{z} + \mathbf{b} \Leftrightarrow (\mathbf{I} - \mathbf{M})\mathbf{z} = \mathbf{b}$$

- $\mathbf{z}$ $(|\widetilde{A}| \times 1)$, $z_k = e^{\frac{1}{\mu}V(k)}$
- $\mathbf{b}$ $(|\widetilde{A}| \times 1)$, $b_k = 0\ \forall k \in A$, $b_k = 1, k = d$
- $\mathbf{M}$ $(|\widetilde{A}| \times |\widetilde{A}|)$

$$M_{ka} = \begin{cases} \delta(a|k)e^{\frac{1}{\mu}v(a|k)} & \forall a \in \widetilde{A},\ \forall k \in A \\ 0 & \text{otherwise} \end{cases}$$

PROPERTIES

- The system has a solution if $(\mathbf{I}-\mathbf{M})$ is invertible which depends on the balance between the number of paths that connect nodes in the network and the size of $\frac{1}{\mu}v(a|k)$
- $\frac{1}{\mu}v(a|k)$ are network dependent since they are scaled to the variance of the error terms
- Dense networks and many alternative paths do not necessarily imply that $\mathbf{I}-\mathbf{M}$ is ill-conditioned

PROPERTIES

- Path probabilities
- Observation $\sigma = \{k_i\}_{i=0}^{I}$, k_0 is the origin and $k_I = d$
- By the Markov property $P(\sigma) = \prod_{i=0}^{I-1} P(k_{i+1}|k_i)$

$$
\begin{aligned}
P(\sigma) &= \prod_{i=0}^{I-1} e^{\frac{1}{\mu}(v(k_{i+1}|k_i)+V(k_{i+1})-V(k_i))} \\
&= e^{-\frac{1}{\mu}V(k_0)} \prod_{i=0}^{I-1} e^{\frac{1}{\mu}v(k_{i+1}|k_i)}
\end{aligned}
$$

- Denote $v(\sigma) = \sum_{i=0}^{I-1} v(k_{i+1}|k_i)$

PROPERTIES

$$P(\sigma) = \frac{e^{\frac{1}{\mu}v(\sigma)}}{e^{\frac{1}{\mu}V(k_0)}} = \frac{e^{\frac{1}{\mu}v(\sigma)}}{\sum_{\sigma' \in \Omega} e^{\frac{1}{\mu}v(\sigma')}}$$

- Path based multinomial logit model with an infinite number of alternatives
- Ω set of all paths (including those with loops)
- IIA property holds over paths

ESTIMATION

- A sample of observations of (different) travellers making path choices σ_n, $n = 1, \ldots, N$, e.g. collected by GPS
- Maximum likelihood estimation

$$\max_{\beta} LL_N(\beta) = \frac{1}{N} \sum_{n=1}^{N} \ln P(\sigma_n; \beta) \qquad (1)$$

ESTIMATION

- Log-likelihood function

$$LL(\beta) = \ln \prod_{n=1}^{N} P(\sigma_n;\beta) = \frac{1}{\mu} \sum_{n=1}^{N} \sum_{t=0}^{T_n-1} v(k_{t+1}|k_t;\beta) - V(k_0;\beta)$$

- Unlike a static logit model we need to solve the value functions in order to evaluate choice probabilities
- A non-linear optimization algorithm is used to search over the parameter space
- We use the NFXP algorithm since the value functions are not expensive to compute

APPLICATION

- Borlänge network (3077 nodes, 7459 links, 21452 link pairs)
- The value functions can be solved using a direct solver (MATLAB)
- Validation using synthetic data
- Estimation based on real observations

APPLICATION – VALIDATION

- Instantaneous utilities

$$v(a|k) = \beta_{TT}TT_a + \beta_{LT}LT_{a|k} + \beta_{LC}LC_a + \beta_{UT}UT_{a|k}$$

- Travel time in minutes (TT)
- Left turn dummy (LT) for turns larger than 40 degrees
- Link constant (LC) to penalize paths with many crossings
- U-turn dummy (UT) to remove u-turns from the network, fixed to -20

APPLICATION – VALIDATION

- One origin-destination pair in Borlänge
- 10 samples of 500 paths observations
- Chosen parameters: $\beta_{TT} = -2$, $\beta_{LT} = -1$, $\beta_{LC} = -1$ and $\beta_{UT} = -20$

APPLICATION – VALIDATION

Sample	$\widehat{\beta}_{TT}$	Std. Err.	$\widehat{\beta}_{LT}$	Std. Err.	$\widehat{\beta}_{LC}$	Std. Err.
1	-1.91	0.21	-1.02	0.09	-1.07	0.06
2	-1.97	0.22	-0.99	0.09	-1.04	0.06
3	-1.80	0.21	-1.09	0.09	-1.07	0.06
4	-2.38	0.26	-0.88	0.09	-1.01	0.06
5	-2.20	0.24	-0.96	0.08	-0.93	0.05
6	-2.30	0.26	-0.96	0.09	-0.96	0.06
7	-1.69	0.18	-1.00	0.08	-1.11	0.06
8	-1.84	0.20	-1.04	0.08	-1.04	0.05
9	-2.40	0.29	-1.05	0.09	-0.89	0.06
10	-1.88	0.20	-0.99	0.08	-0.976	0.05
Average	-2.04	0.23	-1.00	0.09	-1.01	0.06
Std. Err.	0.26		0.06		0.07	

APPLICATION – REAL DATA

- Real data: 1832 path observations, 466 destinations and over 37000 link choices
- Comparison with path based logit models

Parameters	RL-LS	RL	PSL	PL
$\widehat{\beta}_{TT}$	-3.060	-2.494	-2.738	-2.431
Rob. Std. Err.	0.103	0.098	0.086	0.083
Rob. t-test (0)	-27.709	-25.449	-31.837	-29.289
$\widehat{\beta}_{LT}$	-1.057	-0.933	-1.000	-0.920
Rob. Std. Err.	0.029	0.030	0.027	0.029
Rob. t-test (0)	-36.448	-31.100	-37.037	-31.724
$\widehat{\beta}_{LC}$	-0.353	-0.411	-0.545	-0.429
Rob. Std. Err.	0.011	0.013	0.012	0.013
Rob. t-test (0)	-32.091	-31.615	-45.417	-33.000
$\widehat{\beta}_{UT}$	-4.531	-4.459	-4.366	-4.375
Rob. Std. Err.	0.126	0.114	0.118	0.119
Rob. t-test (0)	-35.960	-39.114	-37.000	-36.765
$\widehat{\beta}_{LS}$ and $\widehat{\beta}_{PS}$	-0.227		1.461	
Rob. Std. Err.	0.013		0.082	
Rob. t-test (0)	-17.462		17.817	
$\widehat{LL}_N(\widehat{\beta})$	-3.300	-3.441	-1.601	-1.688

CHAIRE CN INTERMODALITÉ DES TRANSPORTS

CROSS VALIDATION

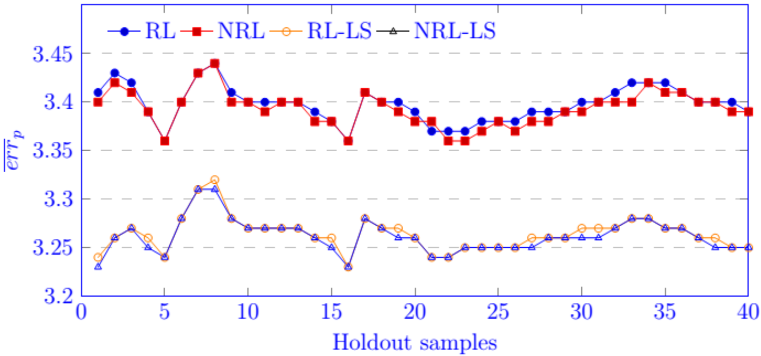
- Cross validation approach: 80% for estimation, 20% for prediction
- 40 holdout samples of the same size taken from the real sample
- Log-likelihood loss, averaged over holdout samples

$$\overline{err}_p = \frac{1}{p}\sum_{i=1}^{p} err_i \quad \forall 1 \leq p \leq 40$$

where err_i is the log-likelihood loss for holdout sample i

CHAIRE CN INTERMODALITÉ DES TRANSPORTS

CROSS VALIDATION



CHAIRE CN INTERMODALITÉ DES TRANSPORTS

BEYOND THE RECURSIVE LOGIT...

- Following Aguirregabiria and Mira (2010)
- Individuals make sequential decisions, are forward looking and maximize expected intertemporal utility
- Parameters to be estimated describe decision makers preferences and beliefs about the future

INTRODUCTION

- We assume that time is discretized $t = 1, 2, \ldots, T$ and the horizon T can be finite of infinite
- A vector of state variables s_{nt} for individual n and time t describes all relevant information for the decision of interest known at t
- n takes an action/choice $a_{nt} \in A_{nt}$

INTRODUCTION

- Instantaneous/current utility of action a_{nt} is $u(a_{nt}, s_{nt}; \beta)$ (time separability is assumed)
- An individual's preferences over time are then represented by

$$\sum_{j=0}^{T} \rho^j u(a_{n,t+j}, s_{n,t+j}; \beta)$$

where $\rho \in (0,1)$ is a discount factor (ρ is assumed constant over time)

INTRODUCTION

- The decision at t affects the values of future state variables and these values are uncertain to the individual
- The evolution of future states is represented by a Markov transition distribution function $F(s_{n,t+1}|a_{nt}, s_{nt})$

INTRODUCTION

- At each time t, an individual observes s_{nt} and chooses $a_{nt} \in A_{nt}$ to maximize expected utility

$$E\left(\sum_{j=0}^{T-t} \rho^j u(a_{n,t+j}, s_{n,t+j}; \beta) | a_{nt}, s_{nt}\right)$$

INTRODUCTION

- Value function recursively defined by Bellman's equation

$$V(s_{nt}) = \max_{a \in A_{nt}} \left[u(a, s_{nt}; \beta) + \rho \int V(s_{n,t+1}) dF(s_{n,t+1} | a, s_{nt}) \right]$$

- Parameters to be estimated can be included in the instantaneous utilities and the transition probabilities in addition to the discount factor

CHAIRE CN INTERMODALITÉ DES TRANSPORTS

DATA

- Panel data of $n = 1, \ldots, N$ individuals
- Actions (choice variable) a_{nt} are observed
- $s_{nt} = (x_{nt}, \varepsilon_{nt})$ where attributes in x_{nt} are observed
- Data $\{a_{nt}, x_{nt} : n = 1, \ldots, N; t = 1, \ldots, T_n\}$

RUST'S MODEL

- Rust (1987) formulated a dynamic discrete choice model which has several similarities with a static logit
- We present these assumptions informally (see e.g. Aguirregabiria and Mira, 2010, for details)

RUST'S MODEL

- Instantaneous utilities are additively separable in observed and unobserved components $u(a, s_{nt}) = v(a, x_{nt}) + \varepsilon_{nt}(a)$
- Random terms $\varepsilon_{nt}(a)$ are i.i.d. Extreme value type I with zero mean and are independent over time, individuals and actions and are also independent of everything else in the model
- $x_{it} \in X$ is discrete and finite
- These assumptions imply that $F(s_{n,t+1}|a_{nt}, s_{nt}) = F_{\varepsilon}(\varepsilon_{n,t+1})F_x(x_{n,t+1}|a_{nt}, x_{nt})$

RUST'S MODEL

The "integrated value function" $\bar{V}(x_{nt}) \equiv \int V(x_{nt}, \varepsilon_{nt}) dF(\varepsilon_{nt})$ is the unique solution to the "integrated Bellman equation"

$$\bar{V}(x_{nt}) = \int \max_{a \in A_{nt}} \left[v(a, x_{nt}; \beta) + \varepsilon_{nt}(a) + \rho \sum_{x_{n,t+1}} \bar{V}(x_{n,t+1}) f_x(x_{n,t+1} | a, x_{nt}) \right] dF(\varepsilon_{nt})$$

To simplify coming formulae, we define a choice specific value function

$$V(a, x_{nt}; \beta) = v(a, x_{nt}; \beta) + \varepsilon_{nt}(a) + \rho \sum_{x_{n,t+1}} \bar{V}(x_{n,t+1}) f_x(x_{n,t+1} | a, x_{nt})$$

RUST'S MODEL

- We now define the probability of choosing an action at a given state
- Individuals maximize instantaneous and expected future utilities

$$\alpha(x_{nt}, \varepsilon_{nt}) = \arg\max_{a \in A_{nt}} V(a, x_{nt}; \beta)$$

- We can obtain choice probabilities as in the static logit case by integrating out the random terms

RUST'S MODEL

- Probability of choosing action a given observed state variables x_{nt}

$$P(a|x_{nt};\beta) = \frac{\exp[V(a,x_{nt};\beta)]}{\sum_{j\in A_{nt}} \exp[V(j,x_{nt};\beta)]}$$

- Given the logit model, the expected maximum utility is the logsum so we obtain a familiar expression for the integrated value function

$$\bar{V}(x_{nt}) = \ln\left(\sum_{j\in A_{nt}} \exp\left[v(a,x_{nt}) + \rho \sum_{x_{i,t+1}} \bar{V}(x_{n,t+1}) f_x(x_{n,t+1}|a,x_{nt})\right]\right)$$

CONCLUDING REMARKS

- In the case of recursive logit the value functions are easy solve because the next state is deterministically given by the action
- When this is not the case, the value functions can be more difficult and time consuming to compute
- NFXP algorithm may be too expensive but alternatives exist (e.g. Hotz and Miller, 1993, Aguirregabiria and Mira, 2002)
- Another lecture focuses on the estimation problem

References

Aguirregabiria and Mira (2002). Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models. Econometrica, 70(4):1519-1543.

Aguirregabiria and Mira (2010). Dynamic discrete choice structural models: A survey. Journal of Econometrics 156:38-67.

Fosgerau, Frejinger and Karlstrom (2013). A link based network route choice model with unrestricted choice set, Transportation Research Part B, 56:70-80.

Hotz and Miller (1993). Conditional choice probabilities and the estimation of dynamic models. The review of economic studies Ltd., 60(3):497-529.

References

Mai, Fosgerau and Frejinger (2015). A nested recursive logit model for route choice analysis,Transportation Research Part B 75(1):100-112.

Rust (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. Econometrica, 55(5):999-1033.

Rust (1994). Structural estimation of Markov decision processes. In: R.F. Engle and D.L. McFadden (Eds.), Handbook of Econometrics, Volume IV, Elsevier Science.