Exam Advanced Econometrics : semi-parametric and simulations

Topic 4: Arcidiacono and Miller (2011)

Maxime Brun, Loïc Cantin and Thibault Ingrand

May 13, 2022

Table of contents

1 Introduction

The expression of an economic problem set in the form of a dynamic discrete choice model is one of the major topics of modern structural econometry. Its solution has long been tedious, sometimes requiring heavy computations to obtain the value function either via backward recursion or fixed-point algorithm. The contribution of Arcidiacono and Miller pursues the efforts initiated by Hotz and Miller (1993) - though their conditional choice probability estimation (CCP) method - by extending the approach. This contribution is essentially twofold, one being the extension of the classes of models to which CCP estimation could be applied and the second being the possibility of considering unobserved heterogeneity depending on various states and possibly varying over time.

The objective of this report is therefore to briefly summarize the method and results of the paper in a first step. We will then provide details on the implementation of the algorithm from a practical point of view and replicate the simulations described in the paper (we focus on the first application - i.e. the bus engine replacement problem). Finally, we propose to apply this model to real data (Rust replacement problem) taking into account unobserved heterogeneity.

2 Summary of the paper

This paper by Arcidiacono and Miller [1] is divided into 8 sections. The first introduces the discussion by reviewing the context of this publication and the main achievements of this paper. Section 2 motivates the development of the method with the example of Rust's bus engine problem. Sections 3 to 5 present respectively the framework of the analysis, the estimation issue, and the details of the algorithm. Section 6 focuses on the estimation of the unobserved heterogeneity parameters. Section 7 presents the results of the application of the method to two problems (engine replacement problem, entry exit problem) performed in Monte Carlo experiments on simulated data and Section 8 concludes.

2.1 Motivation and research objective

This paper is therefore a direct continuation of the Hotz and Miller study and builds upon the CCP approach introduced therein. The advantages of this method are its robustness and flexibility, which means that it does not require the calculation of a new value function for each application. However, if the benefits of introducing the CCP method seem obvious, it did not allow to consider a possible heterogeneity which is unobserved by the econometrician and drives the choices of the agent. The present paper therefore addresses this issue by adjusting the expectation maximization algorithm. The two authors also modify the original algorithm (from the Hotz and Miller paper) by pointing out that the likelihood of a choice can be obtained considering the current payoffs and conditional choice probabilities that occur in the future. The aim of this paper is therefore to propose a relatively simple and flexible estimation method based on an expectation minimization algorithm, while maintaining the theoretical framework provided by a structural model and fitting the data by considering the potential existence of several groups (whose composition is not observed).

2.2 Main findings

The main theoretical results demonstrated by the two co-authors concern the properties of the model they propose and the behaviour of their estimator. Concerning the model - as mentioned earlier - the major contribution is to show that the conditional value function can be written as a combination of the flow payoffs and the conditional choice probabilities for all policies. Another important technical result is to show that the value functions of the different choices do not enter directly into the calculation of the likelihood but that only their difference matters.

Concerning the estimators, it is shown that the estimators $(\hat{\theta}, \hat{\pi}, \hat{p})$ from (θ, π, p) respectively are consistent. θ, π and p represents respectively: the parameter of the coefficients which intervene in the utility function, the transition function and the pdf, the initial distribution of the unobserved heterogeneity, and the conditional choice probabilities. The authors show that it is also possible to obtain an expression for the asymptotic covariance matrix of the estimators (see their supplementary material).

2.3 Description of the algorithm

We present here only a brief summary of the steps of the algorithm (transcribing the summary of part 5.3 of the paper) whose steps will be described in detail in section 3:

- **Step 1**: Find the value of the conditional probabilities of being in each unobserved state (we can consider every finite number of states) $\mathbf{q_{nst}^{(m+1)}}$ using:

$$q_{nst}^{(m+1)} = \frac{L_n^{(m)}(s_{nt} = s)}{L_n^{(m)}}$$

  ,
  where $L_n^{(m)}$ denotes the likelihood of the data on n given the parameters at m_{th} iteration and $L_n^{(m)}(s_{nt} = s)$ denotes the joint likelihood of the data and unobserved state s occurring at time t, given the parameter evaluation at iteration m.

- **Step 2**: Compute the set of initial population probabilities of the unobserved states $\boldsymbol{\pi}^{(\mathbf{m+1})}(\mathbf{s_1}|\mathbf{x_1})$ and the transition parameters for the unobserved states $\boldsymbol{\pi}^{(\mathbf{m+})}(\mathbf{s'}|\mathbf{s})$ using:

$$\pi^{(m+1)}(s|x) = \frac{\sum_{n=1}^{\mathcal{N}} q_{nst}^{(m+1)} I(x_{n1} = x)}{\sum_{n=1}^{\mathcal{N}} I(x_{n1} = x)} \text{ and } \pi^{(m+)}(s'|s) = \frac{\sum_{n=1}^{\mathcal{N}} \sum_{t=2}^{\tau} q_{ns't|s}^{(m+1)} q_{ns,t-1}^{(m+1)}}{\sum_{n=1}^{\mathcal{N}} \sum_{t=2}^{\tau} q_{ns,t-1}^{(m+1)}}$$

- **Step 3**: Compute the conditional choice probabilities $\mathbf{p^{(m+1)}}$ using:

$$p_{jt}(x, s) = \frac{E\left[d_{njt} q_{nst} I(x_{nt} = x)\right]}{E\left[q_{nst} I(x_{nt} = x)\right]} \text{ or } p_{jt}^{(m+1)}(x, s) = \frac{\sum_{n=1}^{\mathcal{N}} d_{njt} q_{nst}^{(m+1)} I(x_{nt} = x)}{\sum_{n=1}^{\mathcal{N}} q_{nst}^{(m+1)} I(x_{nt} = x)}$$

- **Step 4**: Find the structural parameters $\boldsymbol{\theta}^{(\mathbf{m+1})}$ by maximizing:

$$\sum_{n=1}^{\mathcal{N}} \sum_{t=1}^{\tau} \sum_{s=1}^{S} \sum_{j=1}^{J} q_{nst}^{(m+1)} \ln \mathcal{L}_t \left(d_{nt}, x_{n,t+1} | x_{nt}, s_{nt} = s; \boldsymbol{\theta}, \boldsymbol{\pi}^{(\mathbf{m+1})}, \mathbf{p}^{(\mathbf{m+1})}\right)$$

3 Replication of the simulations and estimations

To illustrate the theoretical results of the paper, the authors run Montecarlo simulations of two different discrete choice models and estimate the structural parameters using the algorithm they have developed. We replicate the simulations and results of their first example: the bus engine replacement model.

Rust (1987) gave an example of a discrete choice model, the Bus Engine Replacement Problem, which he tried to estimate the structural parameters of using actual data. Arcidiacono & Miller (2011) extend this simple framework by introducing unobserved heterogeneity, simulate the data of this extended model and use their algorithm to estimate the parameters.

3.1 The Bus Engine Replacement Model

At each period t, Harold Zurcher, the manager of a bus company, decides whether to replace the engine of a bus i. $d_{1it} = 1$ if the engine is replaced, $d_{2it} = 1$ if the engine is kept for at least one more period. $d_{1it} + d_{2it} = 1$. The decision depends on the "brand" of the bus $s_i \in \{1, 2\}$ and on its accumulated mileage x_{1it}. s_i and x_{1it} are observed by Harold, while the econometrician only observes x_{1it}.

Description of the mileage accumulation.

We assume that the accumulation of mileage at period $t+1$ depends on accumulated mileage until time t, x_{1it}, the decision to replace the engine d_{1it+1}, and a *permanent route characteristic* of the bus i, x_{2i}. We also assume that x_{2i} is a multiple of 0.01 and is drawn from a discrete equiprobability distribution between 0.25 and 1.25. Finally, we assume that mileage accumulates in increments of 0.125 and that maintenance costs increase with accumulated mileage up to 25 and then flatten out. The probability of x_{1it+1} conditional on x_{1it}, x_{i2} and d_{jit+1} is specified as:

$$f_j(x_{1it+1}|x_{1it}) = \begin{cases} e^{-x_{2i}(x_{1it+1}-x_{1it})} - e^{-x_{2i}(x_{1it+1}+0.125-x_{1it})} & \text{if } j = 2 \text{ and } x_{i1t+1} \geq x_{i1t} \\ e^{-x_{2i}x_{1it+1}} - e^{-x_{2i}(x_{1it+1}+0.125)} & \text{if } j = 1 \text{ and } x_{i1t+1} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Structure of the payoffs.

We assume that changing the engine induces a fixed-cost θ_0. Then, the payoff of replacing the engine of bus i at time t is $-\theta_0 + \epsilon_{1it}$. Keeping the engine yields a payoff depending on the accumulated mileage and on the brand. Then, the payoff of keeping the engine of bus i at time t is: $\theta_1 x_{1it} + \theta_2 s_i + \epsilon_{2it}$.Therefore, the net payoff of keeping the engine of bus i at time t is $NP_{it} = \theta_0 + \theta_1 x_{1it} + \theta_2 s_i + \epsilon_{2it} - \epsilon_{1it}$.

We assume that the random shocks/unobserved payoffs follow a Type I generalized extreme value distribution.

Decision of replacing the engine.

Zurcher takes into account the current period payoff as well as how his decision today will affect the future, with the per-period discount factor β, i.e. he maximizes the following expected discounted sum of payoffs:

$$EV = E\left(\sum_{t=1}^{+\infty}\beta^{t-1}[d_{2t}(\theta_1 x_{1t}+\theta_2 s+\epsilon_{2t})+d_{1t}(-\theta_0+\epsilon_{1t})]\right)$$

Let $V(x_{1t}, s)$ denote the ex ante value function at the beginning of period t. It is the discounted sum of current and future payoffs just before the random shocks ϵ_{1t} and ϵ_{2t} are realized and before the decision at t is made. We also define the conditional value function for choice j as the current period payoff of choice j net of ϵ_{jt} plus the expected future utility of Harold Zurcher behaving optimally in the future:

$$v_j(x_{1t}, s) := \begin{cases} -\theta_0 + \beta V(0, s) & \text{if } j = 1 \\ \theta_1 x_{1t} + \theta_2 s + \beta V(x_{1t+1}, s) & \text{if } j = 2 \end{cases}$$

Conditional Choice Probability.

We can now express the conditional choice probability (CCP) of replacing the engine of bus i at time t given x_{1it} and s_i.

$$p_1(x_{1it}, s_i) := P(d_{1it} = 1|x_{1it}, s_i) = P(v_2(x_{1it}, s_i) + \epsilon_{2it} > v_1(x_{1it}, s_i) + \epsilon_{1it}))$$

Then, because of the distribution of ϵ_1 and ϵ_2:

$$p_1(x_{1it}, s_i) = \frac{1}{1 + \exp[v_2(x_{1it}, s_i) - v_1(x_{1it}, s_i)]}$$

Finally, the authors show that:

$$\begin{aligned} v_{2it}(x_{1it}, s_i) - v_{1it}(x_{1it}, s_i) = {} & \theta_0 + \theta_1 \min\{x_{1it}, 25\} + \theta_2 s_i \\ & + \beta \sum_{x_{1it+1}} \ln(p_{1t}(x_{1it}, s_i))(f_1(x_{1it+1}|x_{1it}) - f_2(x_{1it+1}|x_{1it})) \end{aligned} \quad (1)$$

Therefore, the probability of replacing the engine $p_1(x, s)$ can be expressed as a function of keeping the engine $\theta_0 + \theta_1 x_1 + \theta_2 s$, the discount factor β, and the one-period-ahead probabilities of replacing the engine.

3.2 Simulation of the Data

The authors simulate the data for a decision-maker living 30 periods and making decisions on 1,000 buses. The simulation procedure is as follows.

Compute value functions by backward recursion for every possible mileage, observed permanent characteristic, unobserved state, and time. Draw permanent observed (x_2) and unobserved (s) characteristics from discrete uniform distributions with support 101 and 2, respectively, and start each bus at zero mileage. Given the parameters of the utility function (θ_0, θ_1, θ_2 and β), the value function, and the permanent observed and unobserved states, compute the probability of a replacement occurring in the first period. Draw from a standard uniform distribution. If the draw is less than the probability of replacement, the decision in the first period is to replace. Otherwise, keep the engine. Conditional on the replacement decision, draw a mileage transition according to the discrete exponential distribution described above. Continuing this way, decisions and mileage transitions are simulated for 30 periods.

3.3 Estimation of the structural parameters

The algorithm described in Section 2 is adapted to the Bus Engine Replacement Problem in order to estimate the structural parameters from the simulated data. More accurately, the goal is to estimate θ_0, θ_1, θ_2 and β.

Table 1

Steps of the algorithm	Instruction in the code
Initial estimation of CCP p_{1it} with a simple logit	`b1 = optim(b1,wlogitd,Y=(y2==0),X=xx,P=PType,` `method="BFGS")$par`
$FVT_{it} = \sum_{x_{1it+1}} \ln(p_{1t}(x_{1it}, s_i))(f_1(x_{1it+1} \mid x_{1it}) - f_2(x_{1it+1} \mid x_{1it}))$	`fvt1 = fvdataRcpp(b1,RX1,tbin,xbin,Zstate,Xstate,` `xtran,N,T,rep(1,N),hetero)`
Start of EM alogorithm	

Computing the likelihood
with current estimation of the parameters
$l\left(d_{nt} \mid x_{nt}, s_{nt}, \theta, p_1^{m+1}\right) = \frac{\exp(\theta_0 + \theta_{11it} + \theta_2 s_i + \beta FVT_{it})}{1 + \exp(\theta_0 + \theta_{11it} + \theta_2 s_i + \beta FVT_{it})}$
	`U1 = cbind(xccp,fvt1)%*%bccp`
	`Like = (y2*exp(U1)+(1-y2))/(1+exp(U1))`
	`Like2=array(Like,c(N,T,2))`
	`base=apply(Like2,c(1,3),prod)`

Initial probabilities π_s
(Annexe B.1.4 of the original paper)
	`intcond_optim=optim(binit,intcond,like=base,`
	`X=intcondX,method="BFGS")`
	`binit = intcond_optim$par`

Conditional probabilities q_{ist} (Equation 2.17 of the original paper)	`PType=intcondP(binit,base,intcondX)`
New estimation of p_{1it}	`b1 = optim(b1,wlogitd,Y=(y2==0),X=xx,P=PType,` `method="BFGS")$par`
Updating FVT$_{it}$	`fvt1=fvdataRcpp(b1,RX1,tbin,xbin,Zstate,Xstate,` `xtran,N,T,rep(1,N),hetero)`
$\max_{\theta} \sum_{n=1}^{\mathcal{N}} \sum_{t=1}^{\tau} \sum_{s=1}^{S} q_{nst}^{(m+1)} \ln l\left(d_{nt} \mid x_{nt}, s_{nt}, \theta, p_1^{m+1}\right)$	`bccp = optim(bccp,wlogit,Y=y2,X=cbind(xccp,fvt1),` `P=PType,method="BFGS")$par`

3.4 Description and interpretation of the results

We replicate columns 3 and 6 of the Table I in Arcidiacono and Miller (2011) [1]. We simulate the model 50 times. The results suggest that we successfully replicated the simulation of the authors.

Table 2: Simulation - s observed

Statistic	θ_0	θ_1	θ_2	β
N	50	50	50	50
Mean	1.986	-0.142	0.969	0.925
St. Dev.	0.011	0.004	0.020	0.023

Table 3: Simulation - s unobserved

Statistic	θ_0	θ_1	θ_2	β
N	50	50	50	50
Mean	2.007	-0.147	0.981	0.897
St. Dev.	0.103	0.011	0.073	0.060

4 An application to (real) Harold Zurcher's decisions

In this section, we propose an application of the method to Harold Zurcher's real data used by Rust (1987) [2]. Henceforth, we refer this data set as Rust's data.

4.1 Methodology

In this section, we explain the issues we faced when moving from simulated data to Rust data.

A first issue was the transition matrices. The transition probabilities that are assumed in the simulations of Arcidiacono and Miller (2011) [1] were not in line with the evolution of the mileages in Rust's data. A plot of the mileage transition, suggested to pick a lower truncated normal law to describe the process. Precisely, the transition probabilities are given by

$$f_j(x_{1it+1}|x_{1it}) = \begin{cases} 0 & \text{if } j = 2 \text{ and } x_{i1t+1} - x_{i1t} < 0 \\ \frac{\phi(\bar{\mu},\bar{\sigma}^2;x_{i1t+1}-x_{i1t})}{1-\Phi(\bar{\mu},\bar{\sigma}^2;0)} & \text{if } j = 2 \text{ and } x_{i1t+1} - x_{i1t} \geq 0 \\ \frac{\phi(\bar{\mu},\bar{\sigma}^2;x_{i1t+1})}{1-\Phi(\bar{\mu},\bar{\sigma}^2;0)} & \text{if } j = 1 \text{ and } x_{i1t+1} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\bar{\mu}$ and $\bar{\sigma}^2$ are the mean and variance of the 'parent' normal distribution. In the data (see Appendix), we found that $\bar{\mu} = 2.7$ and $\bar{\sigma}^2 = 2$.

We created a new `R` function called `xgrid_Rust` to compute the transition matrices. An important difference is that the transition probabilities are no longer dependent on the intensity parameter x_2. As a consequence, the transition probabilities are the same for all buses.

A second issue was due to the unobserved heterogeneity assumption. In the simulated data of Arcidiacono and Miller (2011) [1] there is only two types of heterogeneity. In order to be consistent with the simulated data of the authors, we assumed that the source of heterogeneity in Rust's data came from the engine types. This choice was both convenient and relevant. Convenient because it created two balanced groups and relevant because engine is likely to affect the replacement decision. Therefore, we split the sample into two categories : the buses with engine 6V ($s = 0$) and the buses with engine 8V ($s = 1$). Note that other splits based on purchase year or purchase price could have been studied. We decided to focus only on engine heterogeneity.

A third issue came from data generating process itself. The simulated data of Arcidiacono and Miller (2011) [1] exhibits a high replacement frequency, whereas replacement decisions are very rare in Rust's data. As a consequence, we argue that the simulated data of Arcidiacono and Miller (2011) [1] is particularly convenient to identify and estimate the parameters of interest. However, with Rust's data, the estimation of the model is very sensitive because the objective functions are instable and often tends to infinite values. To improve the estimation, we drop the buses that are not replaced during the periods of estimation.

A fourth issue was due to the estimation of the parameter β. In Rust (1987) [2], beta is not estimated whereas it is estimated in Arcidiacono and Miller (2011) [1]. We decided to present the two cases when β is and is not estimated. When β is not estimated, another `R` function is called for the estimation of the structural parameters. This new function, called `wlogit_nobeta`, is just a slightly modified version of the author's function `wlogit`.

We run the estimation over 100 periods starting from the 5-th period. We only pick the buses that are operated at least 110 periods.

4.2 Results

In table 4 and in table 5 we present the estimation results when β is estimated along with θ_0, θ_1 and θ_2. To begin with, we note that the estimate of β is close to 0, even negative when the engine type is unobserved by the econometrician (nb. the optimization is unconstrained). Theses results suggest that Zurcher is not patient and takes his decision based on the instantaneous outcome. Besides, we observe that whether or not the engine type is observed by the econometrician does not modify the coefficient estimates of θ_0 and θ_1. However, the estimate of θ_2 is sensitive the estimation approach. Yet, we argue that the estimates of θ_2 are relatively low compared to θ_0. In other words, the engine type seems not to be taken into account in the replacement decision. More importantly, we note that the sign of θ_1 is negative which is consistent with the economic intuition. The interpretation is the following : an increase of 1k in the mileage entails a cost of 0.014 or 0.019.

Table 4: Application on Rust's data - engine observed

	θ_0	θ_1	θ_2	β
Estimate	7.480	-0.014	-0.135	0.033

Table 5: Application on Rust's data - engine unobserved

	θ_0	θ_1	θ_2	β
Estimate	7.441	-0.019	0.341	-0.095

Since we find an estimate of β close to 0, we follow Rust (1987) [2] and re-estimate the model assuming $\beta = 0$. The results are shown in table 6 and in table 7. The interpretation of the results is analogous to the previous paragraph. The main findings of our analysis is that the estimates of θ_0 and θ_1 are particularly robust to the estimation approache. In any configuration we find θ_0 in a range between 6.6 and 7.5, and we find θ_1 in a range between -0.019 and -0.014. Though the model and the paremetric assumptions are different, our results could be compared to those obtain by Rust (1987) [2] in table X. Precisely, θ_0 and θ_1 are the counteparts of RC and θ_{11}. In terms of magnitude and depending on the groups observed, Rust (1987) [2] found a mileage parameter bigger than our estimates. This suggests that relative to Rust we temper the importance of the mileage in Zurcher's replacement decision.

Table 6: Application on Rust's data - $\beta = 0$ - engine observed

	θ_0	θ_1	θ_2
Estimate	6.683	-0.015	0.525

Table 7: Application on Rust's data - $\beta = 0$ - engine unobserved

	theta_0	theta_1	theta_2
Estimate	7.017	-0.014	-0.083

5 Conclusion

Conclude here.

References

[1] Peter Arcidiacono and Robert A Miller. "Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity". In: *Econometrica* 79.6 (2011), pp. 1823–1867.

[2] John Rust. "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher". In: *Econometrica: Journal of the Econometric Society* (1987), pp. 999–1033.

A Empirical transition density

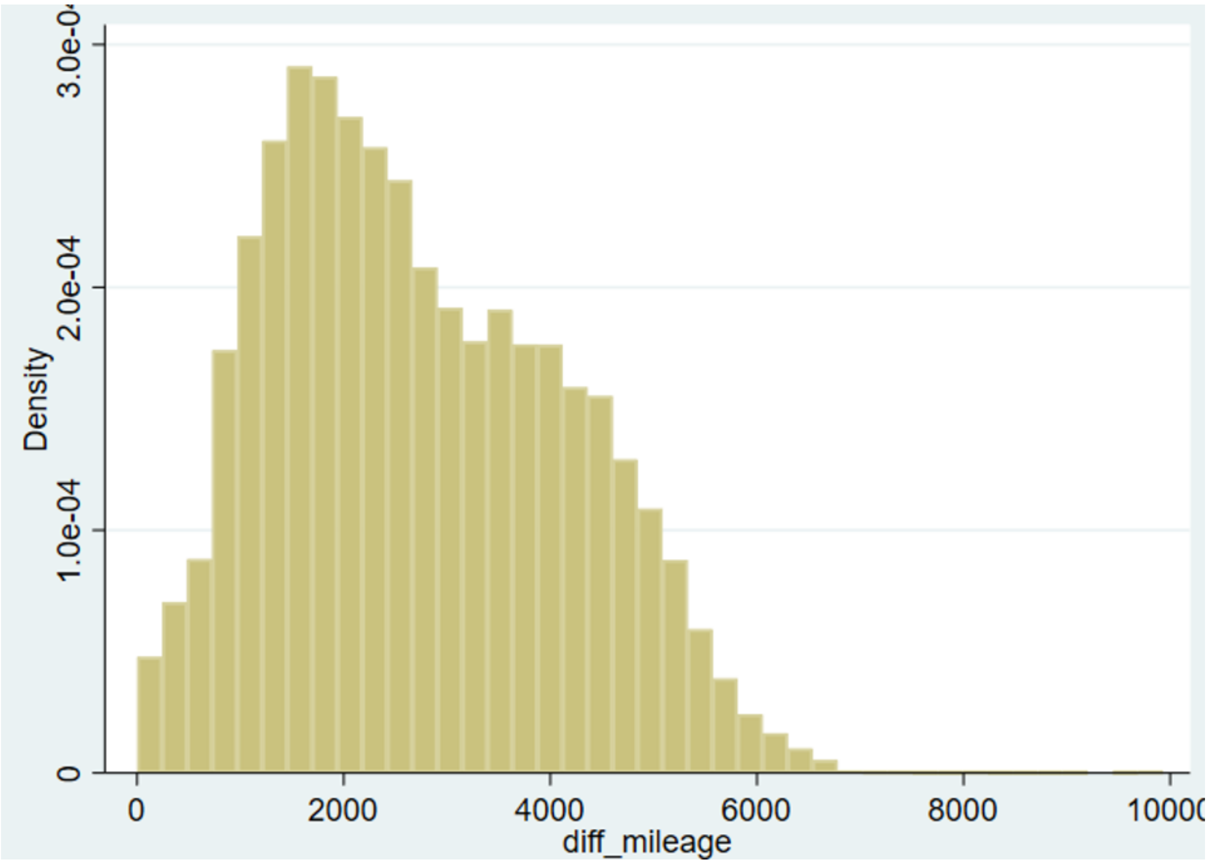

Figure 1: Empirical density of the mileage transition