Car ownership dynamics: An ordered logit approach with irregularly spaced panels

Andrea Papu Carrone[a,*], Mayara Moraes Monteiro[b], Jeppe Rich[a]

[a]*Department of Management Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark*
[b]*Faculdade de Engenharia da Universidade do Porto, R. Dr. Roberto Frias, 4200-465 Porto, Portugal*

Abstract

This paper investigates car ownership dynamics based on irregularly spaced panel data. The data originates from a national travel survey, where a fraction of the respondents are recurrent due to the random sampling scheme of the survey. While this creates a rich sample collected over a long time period and with desired variation between the panel observations, it introduces an estimation challenge due to the irregular spacing of the panel observations. This challenge is addressed by estimating changes in car ownership based on a generalised ordered logit in which the irregular nature of the spacing between panel observations is controlled by including panel-specific weights in the log-likelihood function. The estimated model, which is formulated as a first-difference approach, includes several variables explaining the change of car ownership. Specifically, it is found that accessibility improvements, measured as the number of people that can be reached by public transport within a certain time interval, significantly reduce the likelihood of acquiring additional cars in the observation period. In line with the literature, it is also confirmed that changes in income, number of adults and driver's licenses within the household have a significant impact on household car ownership changes.

Keywords: Panel analysis, Discrete choice, Irregularly spacing, Econometrics, Car ownership dynamics

1. Introduction

Longitudinal data is essential for modelling the relationship between individual's choices (such as car ownership), life course events and exogenous factors that change over time. Yet, studies using long panels are rare (e.g. Prillwitz et al. (2006); Yamamoto (2008)) due to panel attrition (Lugtig (2014)) and the relatively high investment associated with this type of data collection. As a result, researchers typically apply shorter panels comprising 2 or 3 waves (e.g. de Haas et al. (2018); Clark et al. (2016); Scheiner and Holz-Rau (2013)), pseudo panels composed by aggregated cross-sectional data from different years (e.g. Dargay

*Corresponding author email: rich@dtu.dk

and Vythoulkas (1999); Habib et al. (2014)) or retrospective surveys (e.g. Oakil et al. (2014); Fatmi and Habib (2016)). However, while shorter panels are easier to collect they do not constitute a basis for examining the often lagged reactions to life events (Dargay and Vythoulkas (1999)). Pseudo panels, on the other hand, do not have this limitation as these are often based on repeated cross-sectional data. However, in the process of constructing pseudo panels, data is aggregated over individuals with similar characteristics. This leads to information loss and potential aggregation bias in the resulting models (King (1997)). Similarly, retrospective surveys are based on how individuals recall past events and such data is often inaccurate and biased (Oakil et al. (2014)).

An alternative data source, which has not yet been discussed in the literature, is the use of recurrent respondents from national travel surveys. Due to the random sampling scheme of these surveys, it happens that over time respondents will be asked to participate in the survey more than once. In turn, these respondents compose a special type of panel that consists of irregular spaced observations. This alternative data resource has a number of advantages. Firstly, it is a cheap way of getting access to panel data that includes all of the details of a regular cross-sectional travel survey. Secondly, by using irregular spaced panels, one indirectly benefits from heterogeneity in the data as the time-spacing between two consecutive panel observations may vary considerably. While this does mean that the panel observations are scattered over time it also means that the time period covered by the panel is often long. Thirdly, the panel is unlikely to suffer from response fatigue due to the time-spacing. Finally, because irregular spaced panels derived from national travel surveys (mother survey) are essentially created on the basis of randomly selected persons, they are representative with respect to the mother survey and thereby typically also with respect to the population.

2. Data

2.1. Data sources

The analysis presented in the paper is based primarily on data collected by the Danish National Travel Survey. The Danish National Travel Survey (TU, Transportvane-Undersøgelsen in Danish) collects information about the travel behaviour of a representative sample of the Danish population between the age of 10 and 84 years (Christiansen (2018)). The survey is administered by the Center for Transport Analytics of the Technical University of Denmark and has been consistently collected since 2006 interviewing on average 8,200 individuals per year. The participants of the TU are selected by means of a stratified random sample procedure from the Danish Civil Registration System (Det Centrale Personregister, in Danish) in order to form a representative sample of the population.

Although the data collected by the TU is by design cross-sectional, the sample procedure is with 'replacement' which means that, occasionally, individuals will be interviewed more than once. Hence, the longer the survey runs the richer the possibility to derive panel data from the cross-sectional mother survey. The analysis is based on a derived TU panel of 1,547 individuals that have participated in the survey twice but in different years. Since the time-spacing between the two surveys is different for every individual, the paper presents a simple estimation methodology that can be applied when the panel observations are irregularly spaced.

3. Model formulation

3.1. Methods

The model formulation considers the choice of owning a set of cars $c \in C$ at a given point in time $t = 1, \ldots, T$ for a set of households $h = 1, \ldots, H$. The utility function associated with this choice is expressed as:

$$U_{h,t,c} = V_{h,t,c}(x_{h,t,c}) + \epsilon_h + \epsilon_{h,t,c} \tag{1}$$

Where $x_{h,t,c}$ represents a set of explanatory variables that may potentially vary with alternatives $c \in C$, time periods $t \in T$ and households $h \in H$. As presented in Section **??**, it is expected that $x_{h,t,c}$ consists of several confounding variables of which some are stronger predictors for c than others. The first error-term ϵ_h is a panel effect which accounts for variation over households, while the second error term $\epsilon_{h,t,c}$ represents the alternative and time-specific error component for each household.

In a traditional panel context, it is common to translate this problem into first-order differences as a mean to reduce the effect of panel heterogeneity. In this case, we do so by considering the utility function that represents the changes in c for two consecutive time periods, namely $\Delta c = c_t - c_{(t-1)}$. Hence, we consider:

$$\tilde{U}_{h,t}(\Delta c) = \tilde{V}_{h,t,\Delta c}(x_{h,t}) + \tilde{\epsilon}_{h,t,\Delta c} \tag{2}$$

This can reasonably be approximated by:

$$\begin{aligned}\tilde{U}_{h,t}(\Delta c) &= \tilde{V}_{h,t,c}(x_{h,t}) - \tilde{V}_{h,(t-1),c}(x_{h,(t-1)}) + \epsilon_{h,t,c} - \epsilon_{h,(t-1),c} \\ &= \Delta\tilde{V}_{h,t,c}(x_{h,t}) + \Delta\tilde{\epsilon}_{h,t,c}\end{aligned} \tag{3}$$

In the case where $\epsilon_{h,t,c}$ represent IID Gumbel distributed error terms, the difference $\Delta\epsilon_{h,t,c}$ follows a logistic distribution. Hereby follows that, even in the case where the choice of c is modelled as a standard multinomial logit model, the model that explains the differences will belong to the same family of models with logistic distributed errors.

A natural choice of model in this situation is the ordered logit model. One of the shortcomings of the ordered logit model is the assumption of generic attributes across the ordered choices. However, this restriction can be relaxed by considering the generalised ordered logit model or variants of this model.

3.2. Probabilistic model with irregular spaced panels

The above section discusses how a random utility model can be expressed in terms of first-order differences and provides insight into how the choice set and the probability model are affected. However, due to the irregular spaced panels, it is not possible to use a standard estimation technique as the different panel intervals will introduce a scaling problem. Hence, this paper provides an alternative estimator for this problem.

Each individual $n = 1, \dots, N$ is observed at different points in time, $t_0(n)$, ... , $t_T(n)$. In contrast to most common situations, the duration between interviews represented by $\Delta t_t(n) = t_t(n) - t_{(t-1)}(n)$, differs for each individual. That is, a panel data set with irregularly spaced observations.

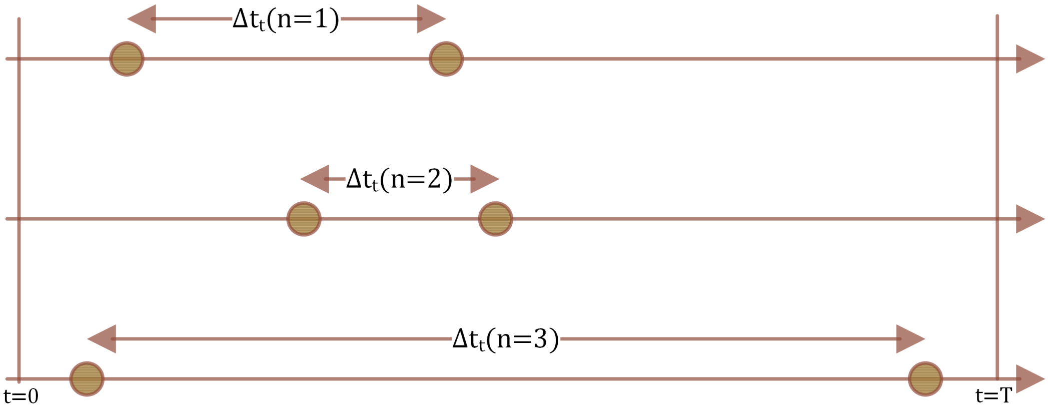

Figure 1: Irregularly spaced panel observations of n individuals.

In order to simplify and tailor the notation to the specific data set, we consider the special case where only two distinct spaced observations exist for each individual as illustrated in Figure 1. We write this as $\Delta t_n = t_t(n) - t_0(n)$ where Δt_n represents the time-spacing between interviews of an individual n. Consider also the corresponding dependent variables $Y_0(n)$ and $Y_t(n)$ and a vector of corresponding explanatory variables $X_0(n)$ and $X_t(n)$. The variables $Y_0(n)$ and $Y_t(n)$ represent the state of a given point in time and the state change for n going from $Y_0(n)$ to $Y_t(n)$ can now be expressed as $\Delta Y_t(n) = Y_t(n) - Y_0(n)$ which reasonably can be modelled as a function of $\Delta X_t(n) = X_t(n) - X_0(n)$.

Clearly, if $\Delta t_n \to \infty$ the probability of a state changing increases all other things equal. If we are to investigate how $\Delta Y_t(n)$ is triggered by changes in life events as captured in $\Delta X_t(n)$, we would ideally like these effects to be measured on a normalised basis. In other words, econometrically speaking, we need to correct for the length of the period. To find the form of this correction term, consider the joint probability of the state space $P_n(\Delta Y_t(n))$ under the assumption of independence over the T regularly spaced intervals between $t = 0$ and t. That is:

$$\begin{aligned} P_n(Y_t - Y_0|X_t, X_0) &= P_n(Y_1 - Y_0|X_1, X_0)P_n(Y_2 - Y_1|X_2, X_1)\dots P_n(Y_t - Y_{(t-1)}|X_t, X_{(t-1)}) \\ &= \prod_{i=1}^{t} P_n(Y_i - Y_{(i-1)}|X_i, X_{(i-1)}) \end{aligned} \tag{4}$$

We can assume that right-hand side probabilities are similar and that each of these expresses an annual rate of change. That is:

$$P_n(Y_t - Y_0|X_t, X_0) = P_n(Y_i - Y_{(i-1)}|X_i, X_{(i-1)})^t \tag{5}$$

From which it follows that:

$$P_n(Y_i - Y_{(i-1)}|X_i, X_{(i-1)}) = \sqrt[t]{P_n(Y_t - Y_0|X_t, X_0)} \tag{6}$$

In other words, when seeking to estimate the unknown probabilistic model for $P_n(Y_i - Y_{i-1}|X_i, X_{i-1})$ this can instead be estimated by considering the empirical weighted log-likelihood function given by:

$$logL(\beta|X_t, X_0) = \sum_{n=1}^{N} \frac{1}{\Delta t_n} ln[P_n(Y_t - Y_0|X_t, X_0)] \tag{7}$$

As a result, the likelihood function reduces to a simple weighted probability model where a change that happens over a long period of time is scaled down to an average annual effect.

4. Model estimation results

4.1. Sample characteristics

As presented in Section 3, the variables are included in the model as differences between the years in which they were observed.

The results of the different models are provided in Table 1. The ordered logit model yields proportional odds and generic parameters estimates (rows indicated with -), while the generalised ordered logit model produces non-proportional estimates (rows indicated with 1 and 0). The final model is a reduced version of the generalised ordered logit, where each of the parameters was tested for proportional and non-proportional odds and removed from the model when insignificant. Therefore, for the final model presented in Table 1, parameters estimates are located in the correspondent row for each of the variables according to the modelling findings. To assess the relative performance of the models, we consider the Log-Likelihood value and Akaike Information Criterion (AIC). As expected, the generalised ordered logit is the model with the highest likelihood, however, when adjusting for the number of additional parameters (as represented by the AIC) and performing a likelihood ratio test the final model is the best performing. Thus, our main attention will be directed to the final model (the model to the right in Table 1). All of the parameters related to variables for socio-economic effects and spatially related features have the expected sign.

Among the socio-economic variables, increases in income and number of driver's licenses in the household positively contribute to the likelihood of increasing the number of cars. This is in line with Prillwitz et al. (2006), that find a positive correlation between increase in the monthly income of a household and car ownership growth. As mentioned in the literature review, increases in the number of driver's licenses are also found to be strong predictors of increases in car ownership levels (Clark et al. (2016)). In regards to the relationship of car ownership levels and changes in employment status, we observe the same patterns as in Prillwitz et al. (2006), namely that becoming employed leads to an increase in household car ownership level. Regarding the influence of household composition, an increase in the number of adults positively affects the likelihood of increasing the number of cars, as previously revealed in the literature (Oakil et al. (2014); de Haas et al. (2018)). Instead, our final model shows that an increase in the number of children in the household does not significantly affect the likelihood to acquire nor relinquish a car. The effect of children in the final model differentiates from the findings of the ordered logit model and the generalised ordered logit model where a positive effect of having children on acquiring a car is observed, as also shown in Yamamoto (2008), Giuliano and Dargay (2006) and Nolan (2010).

Additionally, the analysis of the spatially related features reveals that when the accessibility by public transport increases, the likelihood of acquiring a car significantly decreases. This is in accordance with the findings of Klein and Smart (2019) that considered accessibility to jobs by public transport. Oppositely, when the accessibility by car increases, the likelihood of purchasing a car increases.

Furthermore, there is a very significant and positive effect of gaining home parking facilities on the likelihood to increase the number of cars owned in the household. Changes in the home to work distance do not significantly affect changes in car ownership of a household in our final model, which differs from what Fatmi and Habib (2016) find.

Finally, there is a clear negative effect on the likelihood to acquire a car when relocating to a place with higher population density, which can be seen as a proxy for urbanisation. According to Nolan (2010), this highlights the importance of urban density with respect to household car ownership decisions. Furthermore, taking into account this variable is particularly interesting as a mean to filter out effects that may have otherwise been explained by the accessibility variables.

		Ordered Logit			Gen. Ordered Logit			Final Model		
Variable		Param.	Std Error	p-val	Param.	Std Error	p-val	Param.	Std Error	p-val
Intercept	-									
	1	-1.63	0.08	0.00	0.58	0.13	0.00	-1.66	0.08	0.00
	0	2.19	0.09	0.00	2.12	0.11	0.00	2.29	0.10	0.00
Socio-economic										
Δ*ln(HH income)*	-	0.26	0.12	0.04				0.27	0.12	0.03
	1				0.36	0.22	0.10			
	0				0.05	0.18	0.78			
Δ*D.Lic/members*	-	1.64	0.29	0.00				1.62	0.29	0.00
	1				2.92	0.53	0.00			
	0				1.53	0.40	0.00			
Δ*Numb. adults*	-	0.73	0.09	0.00						
	1				1.20	0.16	0.00	0.50	0.11	0.00
	0				0.83	0.13	0.00	0.92	0.12	0.00
Δ*Numb. children*	-	0.33	0.09	0.00						
	1				0.58	0.17	0.00	0.13	0.11	0.23
	0				0.55	0.13	0.00	0.54	0.12	0.00
Became employed	-	0.52	0.20	0.01						
	1				0.59	0.37	0.12	0.74	0.21	0.00
	0				-0.17	0.34	0.62	0.03	0.32	0.93
Became unemployed	-	-0.16	0.21	0.45						
	1				-0.36	0.36	0.32			
	0				-0.51	0.28	0.07			
Spatially related										
Δ*ln(H-W dist.)*	-	0.17	0.08	0.03						
	1				0.36	0.14	0.01	0.07	0.10	0.48
	0				0.32	0.12	0.01	0.34	0.12	0.00
Δ*Home parking*	-	0.59	0.16	0.00				0.56	0.16	0.00
	1				0.95	0.29	0.00			
	0				0.47	0.23	0.04			
Δ*Dist. to rail*	-	0.19	0.10	0.05						
	1				0.27	0.17	0.12			
	0				0.25	0.14	0.07			
Δ*Pop. density*	-	-0.23	0.06	0.00				-0.24	0.07	0.00
	1				-0.41	0.12	0.00			
	0				-0.13	0.09	0.13			
$\Delta ln(Acc_{PT})$	-	-0.36	0.10	0.00						
	1				-0.82	0.21	0.00	-0.54	0.16	0.00
	0				-0.11	0.10	0.29	-0.32	0.10	0.00
$\Delta ln(Acc_{CAR})$	-	0.41	0.20	0.04				0.37	0.21	0.08
	1				0.69	0.37	0.06			
	0				-0.31	0.29	0.30			
N. of parameters		12			24			15		
AIC		2358.27			2353.84			2346.78		
-2 log L		2330.27			2301.84			2312.78		

Table 1: Model estimates.

5. Conclusion

It is common practise to collect large-scale national travel surveys by sampling individuals randomly and with replacement. As a result, over time a portion of the surveyed individuals will appear in the survey more than once and thereby form a panel of observations. While there will generally not be many recurrent respondents in the panel, such panel data does include information richness by embracing different time-spacing between observations. In some of the cases, there might be only one year between two observations, whereas in other cases, it might be as much as 12 (in this example of the Danish National Travel Survey). Due to this, the panel covers desirable variation in terms of capturing the effect of life events, even if it only captures few observations per individual. Moreover, the survey is cheap as it is indirectly derived from the mother survey and it does not suffer from panel attrition as the spacing between observations prevents this. Another desirable property is that, due to the random sampling, it is commonly representative for the mother survey and thereby for the population.

While there are many desirable properties of such a survey, it does raise a non-trivial estimation challenge in that the time-spacing between observations is not the same. If capturing life events or triggers in the model, it is evident that such triggers are more likely to occur the longer the spacing is. Hence, it is required that we propose an estimation strategy that accounts for this irregular spacing. The paper suggests an individual specific weighting of the likelihood function to overcome this problem, an approach that can easily be extended to all types of discrete choice models. We present the estimation results for three variations of ordered logit models that explain changes in car ownership as a function of changes in a range of variables including income and accessibility. The models are based on approximately 1,500 Danish panel observations and are formulated as a first-order difference approach.

From the final model we conclude that public transport accessibility does have an effect on acquiring more cars in the household even when controlling for population density. This is further explored in a scenario where public transport travel time is changed. In addition, the model includes a range of other confounding variables which all support the findings of the international literature in saying that more driver's licenses, more adults, more children, a new employment and availability of home parking facilities all contribute positively to an increase in the car ownership level of the household.

References

Christiansen, H. (2018). Beregning af usikkerhed og varians for Transportvaneundersøgelsen. *Proceedings from the Annual Transport Conference at Aalborg University*, pages 1–6.

Clark, B., Chatterjee, K., and Melia, S. (2016). Changes in level of household car ownership: the role of life events and spatial context. *Transportation*, 43(4):565–599.

Dargay, J. M. and Vythoulkas, P. C. (1999). Estimation of a dynamic car ownership model: A pseudo-panel approach. *Journal of Transport Economics and Policy*, 33(3):287–302.

de Haas, M. C., Scheepers, C. E., Harms, L. W., and Kroesen, M. (2018). Travel pattern transitions: Applying latent transition analysis within the mobility biographies framework. *Transportation Research Part A: Policy and Practice*, 107(October 2017):140–151.

Fatmi, M. R. and Habib, M. A. (2016). Longitudinal vehicle transaction model: Assessment of lead and lagged effects of longer-term changes and life-cycle events. *Transportation Research Record*, 2566:11–21.

Giuliano, G. and Dargay, J. (2006). Car ownership, travel and land use: A comparison of the US and Great Britain. *Transportation Research Part A: Policy and Practice*, 40(2):106–124.

Habib, K. M., Swait, J., and Salem, S. (2014). Using repeated cross-sectional travel surveys to enhance forecasting robustness: Accounting for changing mode preferences. *Transportation Research Part A: Policy and Practice*, 67:110–126.

King, G. (1997). *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton University Press.

Klein, N. J. and Smart, M. J. (2019). Life events, poverty, and car ownership in the United states: A mobility biography approach. *Journal of Transport and Land Use*, 12(1):395–418.

Lugtig, P. (2014). Panel Attrition: Separating Stayers, Fast Attriters, Gradual Attriters, and Lurkers. *Sociological Methods and Research*, 43(4):699–723.

Nolan, A. (2010). A dynamic analysis of household car ownership. *Transportation Research Part A: Policy and Practice*, 44(6):446–455.

Oakil, A. T. M., Ettema, D., Arentze, T., and Timmermans, H. (2014). Changing household car ownership level and life cycle events: An action in anticipation or an action on occurrence. *Transportation*, 41(4):889–904.

Prillwitz, J., Harms, S., and Lanzendorf, M. (2006). Impact of Life-Course Events on Car Ownership. *Transportation Research Record*, pages 71–77.

Scheiner, J. and Holz-Rau, C. (2013). A comprehensive study of life course, cohort, and period effects on changes in travel mode use. *Transportation Research Part A: Policy and Practice*, 47:167–181.

Yamamoto, T. (2008). The impact of life-course events on vehicle ownership dynamics: The cases of france and Japan. *IATSS Research*, 32(2):34–43.