A dynamic approach to car availability throughout the life course

Romain Crastes dit Sourd   Ilka Dubernet   Matthew Beck
Stephane Hess   Kay W. Axhausen   Christian Holz-Rau
Joachim Scheiner

August 26, 2020

WORKING PAPER - DO NOT QUOTE WITHOUT PERMISSION.

Abstract

The relatively new research field of mobility biographies designates the analysis of long-term mobility behaviour and the availability of mobility tools in a life span. A retrospective survey of the TU Dortmund, ETH Zurich and Goethe University Frankfurt collects data on individual mobility biographies of three different generations in a household with a life-calendar. Most of the past long-term decisions made by individuals such as buying a house or changing job affect their preferences in future periods and induce economic constraints in the form of transaction costs. Ignoring these aspects may lead to biased estimates in the analysis. A dynamic probit model is used to identify impacts on the individual decisions on car availability in a life span and tests for differences between gender or is used to include the time dependency of the explanatory variables such as age, the number of children or education. The focus of the paper is to compare the modelling results following common practices in the life course calendar literature, based on random effects probit models with the results obtained with a dynamic random effects probit model with autocorrelation. In contrary to the classic random effects probit model approach the main advantage of the dynamic probit approach is to explicitly model the correlated time-fixed and time-varying unobserved heterogeneity by the means of composite marginal likelihood estimation.

Keywords: car availability, life course analysis, composite marginal likelihood, mobility biography

1 Introduction and related work

The contemporary increasing complexity of household and family structures, labour markets changes and individualisation of lifestyles cohere with an increase in activities and flexibility, changing attitudes and behaviour patterns. This also affects individual mobility behaviour as well as mobility tool ownership. It is still challenging to capture such ideas conceptually, methodically and empirically, and to identify the most influential factors in order to contribute to planning practice (Axhausen, 2008). So far changes in mobility behaviour are often covered with static cross-sectional studies but these neglect the dynamic and implications of long term decisions (Lanzendorf, 2003).

In the past decade the focus of interest therefore shifted towards individual and joint long term decisions in a life span. The biography approach examines mobility behaviour (including residential choice and travel behaviour) in the context of key events in a life course (such as changes of job or family formation) and life phases (e.g. adolescence or the family phase). Besides people's own experiences the influence of the social environment is of interest. The relevance of both the life course and the social environment are acknowledged in the theoretical discussions about mobility biography and mobility socialisation (Scheiner, 2007). Beige and Axhausen (2012) show that strong interdependencies exist between the various key events and long-term mobility decisions during the life course and argue that events occur to a great extent simultaneously.

Several empirical studies reported in Table 1 attempt to understand and explain everyday travel behaviour as a routine activity changing due to key events such as residential relocation, the birth of a child or exogenous interventions. The different studies are classified based on the data source they use, the type of model they propose (unordered *versus* ordered *versus* continuous), as well as how the influence of past outcomes (the dynamic process) is accounted for in the model. In a large majority of cases, we find that the work based on retrospective data is not accounting for the dynamic nature of the mobility decisions made throughout the life course. A similar conclusion is shared by a comprehensive review of the theoretical framework and most important studies investigating mobility behaviour and mobility tool ownership over the life course recently published by Müggenburg et al. (2015). The authors address open research questions and conclude that studies often investigate long-term decisions with static (panel) models and neglect the dynamic, causality, interrelations and time dependency of the target and explanatory variable (Müggenburg et al., 2015). In addition, most econometric applications hypothesise that car availability (or car ownership) should be modelled as a series of discrete choices following the findings from Bhat and Pulugurta (1998) while recent findings from the qualitative literature indicate that *"car ownership changes are more seen as a process rather than as discrete decisions"* (Clark et al., 2016).

This paper addresses these shortcomings by proposing an autoregressive generalised ordered probit model for modelling car availability, thus taking the earlier described dependencies into account while also engaging with findings from the qualitative literature by representing car availability as a process driven by a single latent variable rather than a choice between several alternatives. Section 2 discusses the factors that have driven car ownership and car availability to be mainly modelled as a succession of discrete choices and proposes an alternative framework. Section 3 further describes the framework of the model, which is subsequently applied to empirical data of a retrospective survey. The paper continues with the description of the data set used for the application in the *Data*

Table 1: Overview on (quantitative) studies on car availability & ownership

Reference	**Data source**			**Model**					**State dependence**			
	Panel	Retrospective	Other	Binomial	Multinomial	Ordered	G-Ordered	Other	Static	Dynamic	Markov	AR
Present study		X				X	X					X
(Beige and Axhausen, 2008)		X		X				X	X			
(Beige and Axhausen, 2012)		X		X					X			
(Cao et al., 2007)			X					X	X			
(Dargay and Vythoulkas, 1999)	X							X		X		X
(Dargay, 2001)	X							X		X		
(Dargay, 2002)	X							X		X		
(Döring et al., 2019)		X		X							X	
(Golob and Van Wissen, 1989)	X					X				X		X
(Golob, 1990)	X					X		X		X		X
(Hanly and Dargay, 2000)	X					X				X		
(Hensher and Le Plastrier, 1985)	X				X					X		
(Hensher, 2013)												
(Kitamura, 1987)	X					X						X
(Kitamura and Bunch, 1990)	X					X				X		
(Mannering, 1983)	X			X					X			
(Meurs, 1993)	X			X	X				X			
(Nobile et al., 1997)	X				X				X			
(Oakil et al., 2014)		X			X				X			
(Prillwitz et al., 2006)	X			X						X		
(Scheiner, 2014)	X							X				X
(Scheiner and Holz-Rau, 2013a)		X						X	X			
(Scheiner and Holz-Rau, 2013b)		X						X				
(Zhang et al., 2014)		X						X		X		

description Section (4). The model results are presented and discussed in the *Results* Section (5). Section 6, *Conclusions and Outlook*, summarizes this paper and gives an outlook on future work and challenges.

2 Ordered and unordered response structures for modelling car availability

In this section, we address a fundamental question which has receive extensive attention in the literature on car ownership and which relates to whether vehicle ownership should be modelled by the means of an ordered or unordered-response mechanism. In a particularly influential paper, Bhat and Pulugurta (1998) present the underlying theoretical structures and identifies the advantages and disadvantages of the two response mechanisms. The unordered response system approach is based on a utility maximisation hypothesis while the ordered response system approach assumes that a single continuous variable represents the latent car ownership propensity of an agent (a household or an individual for example). The authors point out that the unordered response system, which is represented by the well-known multinomial logit model (MNL), is better at capturing a pattern of elasticity effects of variables across alternatives, while the ordered response system, represented by the ordered logit model (OL), is *"constrained to have a more rigid trend in elasticity effects"*.

More precisely, the econometric framework for modelling car ownership *via* a MNL structure can be defined as follows: let the utility of car ownership level k be a function of a vector of exogenous variables x associated with an agent i be written as:

$$U_{k,i} = \beta_k' x_i + \epsilon_{k,i} \tag{1}$$

where β_k is a vector of parameters (including a constant) to be estimated for each car ownership level with k = 1,2,...K. For identification reasons, the parameters for one of the car ownership levels must be normalised to zero. In the notation which we adopt and which is used for modelling *car availability* rather than *car ownership* in the remainder of this paper, all the exogenous variables in the model are associated with agent characteristics as opposed to car features. Following the usual assumption about the error term ϵ being independently and identically distributed across alternative car ownership levels and following a Gumbel distribution, the choice probability for agent i to choose car ownership level k corresponds to:

$$P_{k,i} = \frac{e^{\beta'_k x_i}}{\sum_{k'=0}^{k'=K} e^{\beta'_k x_i}} \tag{2}$$

In contrast, the econometric framework for the ordered response structure, which in this paper will consist in an ordered probit, is defined as follows:

$$y_i^* = \beta' x_i + \epsilon_i, \tag{3}$$

where y_i^* is an unobserved latent process related to *car availability propensity* defined as a function of relevant exogenous variables. i is an index for individuals x_i is a vector of exogenous variables, and β' is a vector of coefficients to be estimated. The error term ϵ_i is assumed to follow a standard normal distribution with zero mean and unit variance (changing this assumption to a Gumbel distribution leads to the OL model used by Bhat and Pulugurta (1998)). The discrete outcome observed for individual i still corresponds to k_i, where k_i may take one value among K. However, we now have that $y_i = k_i$ if $\mu_{k-1,i} < y_i^* < \mu_{k,i}$, where $\mu_{k,i}$ is the upper bound threshold corresponding to the discrete level k_i with $\mu_0 = -\infty$ and $\mu_K = +\infty$. With this notation, μ_1, μ_2, ..., μ_{K-1} are parameters to be estimated with $\mu_1 < \mu_2 < ... < \mu_{K-1}$. The probability for observing a given outcome k for agent i is now given by:

$$P_{k,i} = \Phi(\mu_{k,i} - \beta' x_i) - \Phi(\mu_{k-1,i} - \beta' x_i) \tag{4}$$

where Φ stands for the standard normal cumulative distribution. The ordered structure contrasts with the unordered response structure which features one (latent) utility for each potential outcome, instead of one single latent process related to *car availability*. This important distinction makes the ordered choice model closer to the findings of Clark et al. (2016), who state that "*household car ownership level should be considered as the outcome of a continuous process of development over the life course, rather than as discrete decisions*". However, this does not necessary mean that ordered model structures have been found to be a superior alternative to model car ownership decisions.

Indeed, using four different datasets, Bhat and Pulugurta (1998) estimate a series of models and compare the relative performances of the ordered and unordered response structure by looking at measures of fit in an estimation and a validation sample. Their comparative analysis offers evidence that *"auto ownership modelling must be pursued using the unordered-response class of models"*. However, this is mainly driven by the fact that an unordered model will feature K-1 times more parameters than an ordered model all else being equal. More recent papers have introduced a more flexible class of ordered response structure models known as generalised or hierarchical ordered response models. Such class of models has been originally proposed by Maddala (1986), Ierza (1985) and Srinivasan (2002) with further refinements proposed by Eluru et al. (2008), among others.

2.1 Generalised ordered structures

Generalised ordered response structure models are built on the same principles as the simple ordered structure previously described (which means that such models still feature $K-1$ thresholds), but allow the thresholds mu to vary across agents based on observed or unobserved heterogeneity. In this paper, we are particularly interested in the structure where thresholds are a function of exogenous variables related to the characteristics of the agents, labelled z and which is not restricted to the variables featured in x although this is the case in our application. A similar model is proposed by Eluru (2013) and we adopt an analogous notation:

$$\mu_{k,i} = \mu_{k-1,i} + e^{\kappa_k + \zeta_k' z_i} \tag{5}$$

For identification reasons, one of the thresholds must remain constant across agents, meaning that $\mu_{k,i} = \mu_{k-1,i} + e^{\kappa_k}$. The choice of which threshold to set as the base (or reference) is arbitrary. Assuming a (standard) normally distributed error term for the unobserved latent outcome Y^* leads to the generalised ordered probit model, for which the likelihood function simply corresponds to Equation 8. In the literature, several studies have compared the generalised ordered response structure with an unordered response structure (Eluru, 2013). The generalised ordered response structure, by allowing more flexible thresholds than classic ordered response models, is found to be *on par* with the unordered response structure, which is mainly driven by the fact that such specifications can now feature the exact same number of parameters all else being equal. The choice of an ordered or unordered specification should hence be made based on whether researchers seek to explicitly recognise the inherent ordering within the decision variable or not, as well as whether no major discrepancies can be reported in terms of goodness-of-fit between a generalised ordered response structure and an equivalent unordered model. Having clarified the fact that an unordered response model should not be considered as *a priori* superior to an ordered response structure for modelling car availability and having demonstrated the compatibility of this approach with the qualitative findings of Clark et al. (2016), we now move on to the description of panel generalised ordered response structures for large time series, which are typically encountered in mobility biography studies.

3 Modelling work

3.1 Panel probit

The panel generalised ordered probit model simply consists in introducing a random effect in the cross-sectional generalised ordered model previously introduced. The *latent car availability propensity* for an agent i is given by:

$$y^*_{ij} = \beta' x_{ij} + \alpha_i + \epsilon_{ij}, \tag{6}$$

where j is an index for the j^{th} observation for agent i, with $j = 1, 2, ..., J$ the number of periods under study, x_{ij} is a vector of exogenous variables which can vary across time, and β' is a vector of coefficients to be estimated. The new parameter, α_i, corresponds to an individual specific random disturbance (*i.e.* a random effect). Finally, the serially independent error term ϵ_{ij} is assumed to follow a standard normal distribution with zero mean and unit variance.

The discrete outcome observed for individual i at time j corresponds to k_{ij}, where k_{ij} may take one value among K at each time period ($k_{ij} = 1, 2, ..., K$). In the context of this paper, k_{ij} refers to a given level of car availability among K (*"never available"*, *"sometimes available"* or *"always available"*). We have again that $y_{ij} = k_{ij}$ if $\mu_{ijk-1} < y^*_{ij} < \mu_{ijk}$, where μ_{ijk} is the upper bound threshold corresponding to the discrete level k_{ij} with $\mu_0 = -\infty$ and $\mu_K = +\infty$. We have again that the thresholds are allowed to vary based on observed characteristics of the agents or their environment:

$$\mu_{k,ij} = \mu_{k-1,ij} + e^{\zeta_k + \beta'_k x_{ij}} \tag{7}$$

Finally, $\alpha_i = \alpha + \eta_i$ where η_i is an individual-specific random term. The role of η_i is to generate an equi-correlation between the repeated choice situations for a given individual. The α parameter is normalised to 0 if μ_1 is estimated (and the reverse is also possible). In this paper, we consider that η_i is normally distributed with variance σ^2 but other distributional assumptions may be tested. The model is easily and rapidly estimated using Maximum Simulated Likelihood (MSL). The probability of the observed vector k_i of the sequence of ordinal choices ($k_{i1}, k_{i2}, ..., k_{iJ}$) for individual i given the individual specific random term η_i can be written as:

$$P(k_i)|\eta_i = \prod_{j=1}^{J} \left(\Phi(\mu_{ijk} - \alpha - \beta' X_{ijk-1} - \eta_i) - \Phi(\mu_{ij} - \alpha - \beta' X_{ij} - \eta_i) \right) \tag{8}$$

where Φ stands for the standard normal cumulative distribution. It is then easy to integrate out the individual specific random-term η_i in order to obtain the unconditional log-likelihood of the observed choice sequence.

$$logL_i(\theta) = \\ log\Big[\int_{-\infty}^{+\infty}\prod_{j=1}^{J}\Big(\Phi(\mu_{ijk}-\alpha-\beta'X_{ij}-\sigma v)-\Phi(\mu_{ijk-1}-\alpha-\beta'X_{ij}-\sigma v)\Big)\phi(v)dv\Big] \quad (9)$$

where $v = \frac{\eta_i}{\sigma}$ with $\eta_q \sim N(0, \sigma^2)$ and θ corresponds to a vector of parameters. The log-likelihood function of the *P-GOPROBIT* entails only a one dimensional integral so model estimation is generally fast.

3.2 Autoregressive structures

The simple model presented above assumes that the multiple observations for each individual are equally correlated across time. However, we seek to model *car availability* as a dynamic process, that is to account for the fact that Y_* is influenced by its past realisations. From a behavioural point of view, this means that we assume that some of the unobserved factors affecting *car availability* at time t are correlated with the same unobserved factors at time $t-1, -2, -T$, giving rise to an autoregressive process. We argue that such a modelling approach is necessary to better explain *car availability* changes over time as a process (Clark et al., 2016) in the sense that it better captures the temporal dimension of such process.

We follow Paleti and Bhat (2013) and assume a classic autoregressive structure of order 1 (AR1). We define $corr(\epsilon_{ij}, \epsilon_{ig} = \rho^{|t_{ij}-t_{ig}|})$ with t_{ij} the measurement time for observation y_{ij} ($g \neq j$), where $0 < \rho < 1$, a constraint that be easily enforced through a logistic transformation. The latent outcomes $y*_{ij}$ now follow a multivariate normal distribution for the i^{th} individual. The mean vector of the multivariate normal distribution may be standardised in which case it corresponds to $\frac{\alpha+\beta'X_{i1}}{\tau}, \frac{\alpha+\beta'X_{i2}}{\tau}, ..., \frac{\alpha+\beta'X_{iJ}}{\tau}$ while the correlation matrix Σ has non diagonal entries $\zeta_{ig} = \frac{\sigma^2+\rho^{|t_{ij}-t_{ig}|}}{\tau^2}$, where τ, the standard deviation of the latent outcome $y*_{ij}$, corresponds to $\sqrt{\sigma^2+1}$. While (9) only entails a one-dimension integral, the autoregressive model requires the evaluation of an integral of dimension J for agent i. The log-likelihood function becomes:

$$logL_i(\theta) = \Big[\int_{w_1=\delta_{m_{i1}-1}}^{\delta_{m_{i1}}}, ..., \int_{w_J=\delta^{m_{iJ}-1}}^{\delta_{m_{iJ}}} \phi_J(w_1, ..., w_J|\Sigma)dw_1, ..., dw_J\Big], \quad (10)$$

where $\delta_{m_{ij}} = \frac{\mu^{m_{ij}}-\alpha-\beta'X_{ij}}{\tau}$ and ϕ_J is the standard multivariate normal distribution of dimension J and $w_1, w_2, ..., w_J$ are the normalised means.

The dimensionality of integration often rules out the use of MSL for estimating the autoregressive panel generalised ordered probit model. For example, in the context of the application presented in this paper, such a model would take weeks if not months to converge and would be very prone to simulation errors (Paleti and Bhat, 2013). These issues are easily circumvented by the Composite Marginal Likelihood (CML) estimation approach, which, in the context of this paper, entails only the evaluation of pairs of bivariate normal probabilities.

3.3 Composite Marginal Likelihood estimation

Recently, the Composite Marginal Likelihood (CML) and its developments such as the Maximum Approximate Composite Marginal Likelihood (MACML) methods have become popular alternatives to MSL in the choice modelling field (Bhat, 2011; Varin, 2008; Varin and Czado, 2009; Varin et al., 2011). Composite Likelihood is an inference function derived by multiplying a collection of component likelihoods, where the collection used is determined by the context (Varin et al., 2011), and where each individual component is a conditional or marginal density, leading to an unbiased estimator. Paleti and Bhat (2013) simply describe CML as an estimation technique which replaces the multivariate probability of the dependent choices in the likelihood function by a compounding of probabilities of lower dimensions. A growing literature on CML estimation has proven that this estimation technique can perform as well as MSL at a fraction of the computational cost (Paleti and Bhat, 2013). It is worth noting that there exists a Bayesian approach to CML estimation (Pauli et al., 2011).

3.4 The composite marginal likelihood autoregressive panel ordered probit model

The CML functions presented in this paper are pairwise-likelihood functions formed by the product of likelihood contributions of varying subsets of pairs of observed events. The following equation assumes that all the possible pairs are used for each individual. A typical full-pairwise log-likelihood function for the i^{th} individual corresponds to:

$$logL_i(\theta) = \sum_{g=j+1}^{J} \sum_{j=1}^{J-1} w_i \cdot log\Big[Pr(y_{ij} = k_{ij}, y_{ig} = k_{ig})\Big], \tag{11}$$

where w_i is a weight which varies across individuals in unbalanced panel data contexts and

$$\begin{aligned} &Pr(y_{ij} = k_{ij}, y_{ig} = k_{ig}) \\ &= \phi_2(\delta_{k_{ij}}, \delta_{k_{ig}}, \zeta_{ig}) - \phi_2(\delta_{k_{ij}}, \delta_{k_{ig-1}}, \zeta_{ig}) \\ &- \phi_2(\delta_{k_{ij-1}}, \delta_{k_{ig}}, \zeta_{ig}) + \phi_2(\delta_{k_{ij-1}}, \delta_{k_{ig-1}}, \zeta_{ig}) \end{aligned} \tag{12}$$

It is worth noting that (12) can be evaluated rapidly by using the rectangle properties of the bivariate normal distribution. Varin and Czado (2009) indicate that the CML estimator is consistent and asymptotically normally distributed, where the asymptotic variance covariance matrix is given by the Godambe sandwich information matrix (Godambe, 1960; Zhao and Joe, 2005). The CML formulation is remarkably short and simple in comparison to its MSL counterpart. However, Paleti and Bhat (2013) as well as Varin and Czado (2009), among others, have proved that it is as able as the MSL approach to estimate the model parameters while being less prone to convergence issues. It is important to mention that although 11 uses all possible pairs of observations for each individual. However, the whole set of pairs for each individual may not be necessarily used in practice.

3.5 Model composition

The full-pairwise marginal likelihood function also presented in equation (11) requires the evaluation of $J \times (J-1)/2$ pairs of bivariate normal probabilities in the case of J time periods observed for each individual in the dataset. A full-pairwise approach is efficient and computationally affordable when the number of time periods is moderate, where the definition of moderate depends on the context and the sample size, amongst other factors. The full-pairwise approach becomes more computationally intensive as the number of time periods increases. A recent stream of studies proved that there may be no need to make use of all possible pairs, as pairs formed from closer observations provide more information than distant pairs. This has been found to be true in both temporal and spatial contexts (Bhat et al., 2014; Varin and Vidoni, 2005). Bhat et al. (2014) suggests that the optimal maximum distance d between pairs can correspond to the value that minimises the trace (or the determinant) of the asymptotic variance-covariance matrix of a model with as complete a specification of covariate effects as possible. A similar proposition has been made by Varin and Vidoni (2005). Bhat et al. (2014) and Varin and Vidoni (2005) simply suggest starting with a low value of the distance threshold (which requires the evaluation of a small number of pairs in the CML function) and increase the distance threshold up to a point where increasing it does not improve the trace, or even increases it. Formally, this approach can be defined as the weighted sum of the log-bivariate densities of pairs of observations that are distant apart up to lag d. In this paper, we build our model by following this approach.

Finally, in the context of unbalanced panel data, that is when the number of observations for each individual varies across the sample, it is necessary for CML estimation to be efficient to vary the weight factor w across respondents. In their review, Paleti and Bhat (2013) report a number of weighting strategies and in particular the recent contribution from Joe and Lee (2009), who suggest to set $w_q = (J_i - 1)^{-1}[1 + 0.5(J_i - 1)]^{-1}$. In this paper, we follow the recommendations from Joe and Lee (2009) given that the panel we use is unbalanced, as introduced in the next section.

4 Data description

The data originates from a retrospective survey which is carried out since 2007 at a the Department of Transport Planning of the TU Dortmund as an annual first-year seminar's homework. The questionnaire for the survey was primarily designed as part of a diploma thesis (Klöpper and Weber, 2007) and has been used since then without adjustments to guarantee the comparability of the data. Since 2012 it is part of the collaborative project *"Mobility Biographies: A Life-Course Approach to Travel Behaviour and Residential Choice"* and data is additionally collected in Frankfurt and Zurich.

The survey addresses the students of the seminar, their parents and grandparents. The students represent the seeds and are asked to give the questionnaire to both their parents and two of their grandparents - who are randomly chosen, one from the maternal and one from the paternal side. If one of the family members is not available for any reason the students can alternatively ask another person preferably of the same generation. The questionnaire which is the same for every generation asks for retrospective information on an individual's residential and employment biography, travel behaviour and holiday trips as well as socio-economic characteristics and behavioural attitudes.

From 2007 - 2012 the participation in the survey was mandatory for the students in Dortmund which hence resulted in an average response rate above 90%. In 2013 the students could participate voluntarily thus the rate dropped to almost 20% which is slightly higher but still comparable to the response rates experienced in Frankfurt and Zurich in 2013 where participation was also voluntary. Consequently since 2014 the data collection is again mandatory in Dortmund and also in Frankfurt. Due to university ethical guidelines participation in Zurich remains voluntary. The data is collected at a person level so that every individual represents one case in the dataset. It is also possible to identify the members of one family and model aggregated groups.

4.1 Data issues

As the sample has a unique structure it is not possible to appraise representativeness (see (Erickson, 1979) for problems with representativeness in snowball surveys). The seeds are participants of a university seminar thus due to survey design highly educated individuals are likely to be overrepresented in all three generations. The majority of the respondents live in Dortmund respectively North Rhine-Westphalia - one of the most densely populated regions of Germany - so the data might also contain a bias to a more urban population. Furthermore within the grandparent generation a bias to female participants who live longer on the one hand and are also often younger, more popular and communicative can be recognized (Scheiner et al., 2014). Finally retrospective data especially collected for a long period as the life course always bears the risk of the so called memory bias which means a unintended or voluntary bias of the autobiographic memory (Manzoni et al., 2010). However the whole study focusses on mobility behaviour in the life-course and on finding intergenerational relations thus the results are not expected to be significantly affected by the structural differences between the sample and the population. A more detailed documentation of the data set can be found in Scheiner et al. (2014)

4.2 Survey sample

For this paper data gathered for the parents generation in Dortmund from 2007-2012 is analysed. The dataset contains 684 respondents. The minimum age is 18 years old while the maximum is 72. One observation correspond to one year for a given respondent. The data set features 20044 observations in total, meaning that it is unbalanced. The maximum number of observation for a respondent is 51. The variables which are used for modelling car availability in the remainder of the paper are introduced below:

- *Car availability* (Categorical dependent variable): whether a car is *0 - Never available*, *1 - Sometimes available* or *2 - Always available.*
- *age*: age in years
- *distw*: distance to work, in kilometers
- *german*: 1 if the nationality of the respondent is german, 0 else
- *own_home*: 1 if the respondent owns its home, 0 else
- *moped*: 1 if the respondent owns a moped, 0 else
- *moving*: 1 if the respondent changed home on a given year, 0 else
- *children*: Number of children
- *birth*: 1 if birth of a child on a given year, 0 else
- *married*: 1 if the respondent is married, 0 else
- *wed*: 1 if year of wedding, 0 else
- *licence_moped*: 1 if the respondent has a license for driving a moped, 0 else
- *licence_car*: 1 if the respondent has a license for driving a car, 0 else
- *central 1 t0 5*: whether the respondent lives in a central location or not. 1 if the respondent corresponds to the given category, 0 else. Ranges from very central (central1) to very remote (central5)
- *pop 1 to 7*: whether the respondent lives in a large city of not. 1 if the respondent corresponds to the given category, 0 else. Ranges from less than 1,000 inhabitants (population1) to more than 500,000 inhabitants (population7)

- *edu_cat*: Types of education. 1 if the respondent corresponds to the given category, 0 else.
- *degree*: 1 if the respondent has a degree and 0 else
- *home_cat*: Type of home. 1 if the respondent corresponds to the given category, 0 else.

In addition, we provide more details about the distribution of car availability across time in Figure 1 below:

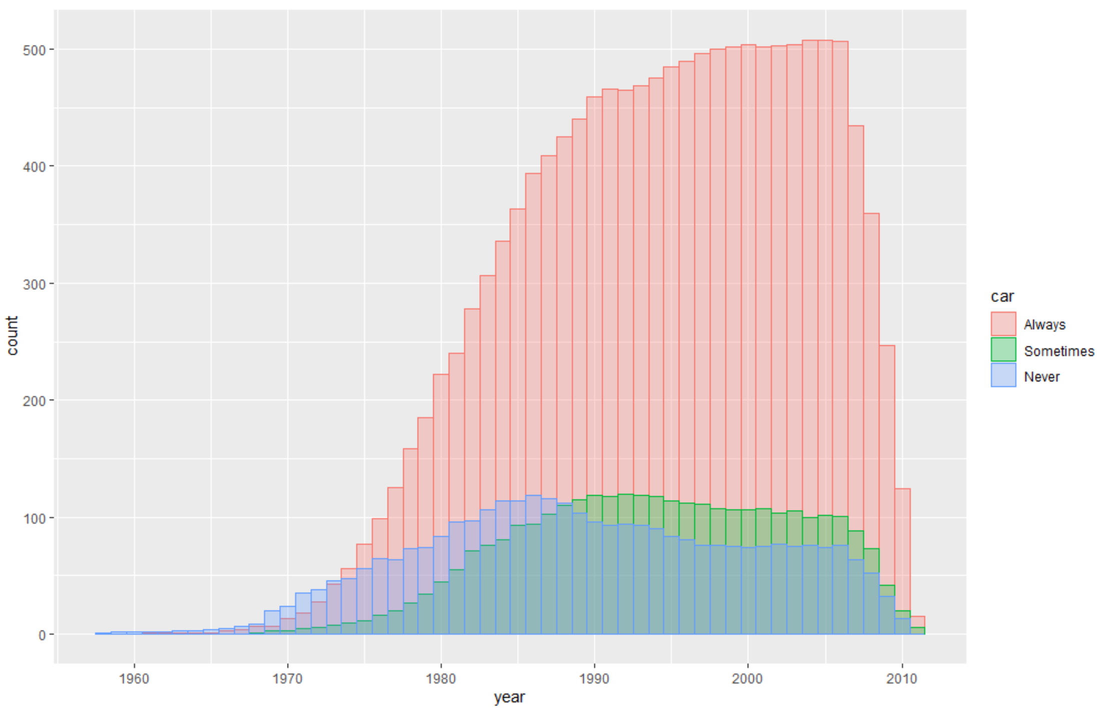

Figure 1: Car availability across time

4.3 Modelling strategy

The objective of this modelling work is to model *car availability* as a dynamic process rather than a series of decisions over time, which calls for using an ordered response structure model while also accounting for the panel nature of the data and autocorrelation in the errors. However, a series of conditions need to be fulfilled before estimating such a model. Firstly, it is necessary to control whether there is an important discrepancy in terms of goodness-of-fit between an ordered and unordered response structure for modelling *car availability*. Secondly, the impact of accounting for the dynamic nature of the process by modelling autocorrelation in the errors needs to be carefully assessed by also estimating a non-dynamic model and comparing results. Finally, the model outputs need to be reported as both parameter estimates and marginal (or partial effects, that is the effect of a discrete change in a given independent variable on the probability of reporting a given level of *car availability* all else being equal) in order to be able to fully understand which factors influence *car availability* through the life course. Five models are estimated in total:

- Model A: cross-sectional ordered probit model
- Model B: cross-sectional multinomial logit model
- Model C: cross-sectional generalised ordered probit model
- Model D: Random effects generalised ordered probit model
- Model E: Autoregressive (AR1) generalised ordered probit model

For the multinomial model, the base is set as *Never available.* For the generalised ordered models, which feature two thresholds parameters μ_1 and μ_2 given that the dependent variable has three modalities, we introduce some flexibility in μ_2 by making it a function of the same variables which are set to affect Y^*, the latent car availability propensity. The model results are reported in the next section.

5 Results

5.1 Unordered *versus* ordered response structures

We being our quantitative analysis by comparing the outputs from Model A, B and C. Model results are reported in Table 2. We do not compare parameter values in details are they are not directly comparable and our interest here is mainly to compare the fit of the ordered and unordered response structures. The log-likelihood for Model A is found to be -13963.03 while it is -13318.89 for Model B and -13328.71 for Model C. The difference between Model B and C is negligible given the magnitude of the log-likelihood for both models. These results indicate that the ordered model structure is as good as the unordered response structure for modelling *car availability* providing that the two models are allowed to be as flexible as one another. We conclude that an ordered response structure is suitable given the data at hand and pursue our analysis with estimating and commenting the results for Models D and E.

Table 2: Ordered and unordered model results

Model A - ordered probit

Parameters	*Latent car availability propensity*	
Log-Likelihood	-13963.03	
AIC	27994.05	
BIC	28262.84	
	Coefficient	Robust T.
Threshold1	1.1501	3.6577
Threshold2	-0.4169	-5.6711
female	-0.6877	-7.0583
age	0.0152	3.2064
distw	0.0137	3.4492
german	0.1682	0.7747
own_home	0.2220	2.3035
moped	0.0337	0.1407
moving	-0.0283	-0.8523
children	-0.0829	-1.7065
birth	0.0790	2.4066
married	0.1899	2.0804
wed	-0.0200	-0.3248
license_moped	0.0970	0.8947
license_car	1.3744	8.2571
central2	0.1185	0.9360
central3	0.3361	2.8136
central4	0.1238	0.9600
central5	0.1909	1.5968
pop1	-0.2518	-0.8399
pop2	0.1979	1.0793
pop3	-0.0124	-0.0809
pop5	-0.0888	-0.5857
pop6	-0.0751	-0.5819
pop7	-0.1796	-1.2710
edu_training	0.2726	2.0521
edu_uni	0.2855	1.9386
edu_other	0.4502	2.8662
degree	0.2067	1.8024
home_det_house	0.0190	0.1561
home_semi_det_house	-0.0934	-0.6780
home_townhouse	-0.1183	-0.9307
home_aprt_big	-0.0887	-0.5630
home_other	0.0904	0.5608

Model B - Multinomial logit

Parameters	*Sometimes*		*Always*	
Log-Likelihood	-13318.89			
AIC	26769.79			
BIC	27291.56			
	Coefficient	Robust T.	Coefficient	Robust T.
.	.	.	.	.
Constant	-4.2598	-4.53	-3.145	-4.57
female	-0.3712	-1.28	-1.4092	-6.17
age	-0.0024	-0.16	0.0262	1.98
distw	0.018	1.36	0.0413	3.91
german	0.5858	1.06	0.4472	1
own_home	0.9619	3.12	0.7678	3.1
moped	-0.5626	-0.88	-0.0253	-0.05
moving	0.0094	0.1	-0.0317	-0.4
children	-0.0214	-0.14	-0.1503	-1.12
birth	0.0644	0.67	0.1568	1.84
married	1.0017	3.6	0.7026	3.15
wed	-0.3549	-1.81	-0.2182	-1.28
license_moped	0.128	0.38	0.2721	0.93
license_car	2.6072	6.46	2.7173	9.19
central2	0.3777	1.01	0.344	1.11
central3	-0.0335	-0.09	0.6302	2.36
central4	0.5186	1.3	0.3718	1.16
central5	0.1466	0.4	0.3714	1.31
pop1	-2.0436	-2.16	-1.0771	-1.74
pop2	-0.5151	-0.89	0.1539	0.36
pop3	-0.028	-0.06	-0.1355	-0.34
pop5	0.1931	0.41	-0.1289	-0.32
pop6	0.0565	0.15	-0.128	-0.4
pop7	-0.1786	-0.43	-0.3992	-1.18
edu_training	0.2466	0.65	0.6794	2.2
edu_uni	-0.4029	-0.96	0.4379	1.34
edu_other	0.7578	1.67	1.2127	3.24
degree	0.7445	2.43	0.4195	1.75
home_det_house	-0.4629	-1.23	-0.0977	-0.32
home_semi_det_house	-0.8185	-2.01	-0.4432	-1.41
home_townhouse	-0.3437	-0.81	-0.3133	-0.9
home_aprt_big	0.3458	0.81	-0.0831	-0.23
home_other	-0.8945	-2.02	-0.0464	-0.14

Model C - Generalized ordered probit

Parameters	*Latent car availability propensity*		*Thresholds and interactions with threshold 2*	
Log-Likelihood	-13328.71			
AIC	26789.41			
BIC	27311.19			
	Coefficient	Robust T.	Coefficient	Robust T.
Threshold1	.	.	1.5216	4.08
Threshold2	.	.	-2.4071	-3.62
female	-0.6359	-5.55	0.1207	0.76
age	0.0109	1.61	-0.0073	-0.88
distw	0.0107	2.44	-0.0051	-0.5
german	0.1762	0.67	0.158	0.36
own_home	0.5005	3.83	0.498	3.05
moped	-0.0556	-0.22	-0.4436	-0.96
moving	-0.0248	-0.63	-0.0016	-0.03
children	-0.0763	-1.16	0.0236	0.3
birth	0.0759	1.76	-0.0051	-0.11
married	0.3932	3.49	0.5461	3.3
wed	-0.1193	-1.42	-0.1696	-1.65
license_moped	0.0424	0.31	-0.0684	-0.4
license_car	1.5565	9.88	1.0469	2.99
central2	0.2041	1.22	0.1639	0.79
central3	0.2269	1.6	-0.256	-1.17
central4	0.2284	1.35	0.2266	1.03
central5	0.1898	1.26	-0.0028	-0.01
pop1	-0.6312	-2.06	-1.1608	-1.62
pop2	0.0449	0.21	-0.3655	-1.03
pop3	0.0257	0.13	0.0979	0.38
pop5	0.0064	0.03	0.1827	0.76
pop6	0.0138	0.08	0.1445	0.66
pop7	-0.1323	-0.74	0.076	0.29
edu_training	0.3016	1.83	0.0328	0.15
edu_uni	0.1686	0.97	-0.3314	-1.27
edu_other	0.6029	2.98	0.2375	0.9
degree	0.3523	2.77	0.5091	2.58
home_det_house	-0.1249	-0.79	-0.3018	-1.46
home_semi_det_house	-0.3473	-2.12	-0.4912	-2.09
home_townhouse	-0.1967	-1.09	-0.1703	-0.81
home_aprt_big	-0.0186	-0.1	0.1409	0.59
home_other	-0.0898	-0.5	-0.6504	-2.27

5.2 Static *versus* dynamic model structures

The estimation of Model D is straightforward. It simply consists in re-estimating Model C while accounting for the panel nature of the data by adding a random effect parameter, σ. The model is estimated using MSL and 1,000 Halton draws. On the other hand and as previously introduced in Section 3.5, finding the best autoregressive model using CML requires to estimate a series of models with an increasing number of related pairs of observations for the same individual. We start by estimating a model where each bivariate pair is not separated by more than two years and increase this number until the logarithm of the determinant of the robust variance-covariance matrix for each estimated model stops increasing (or decreases). Results are reported in Table 3 and Figure 2.

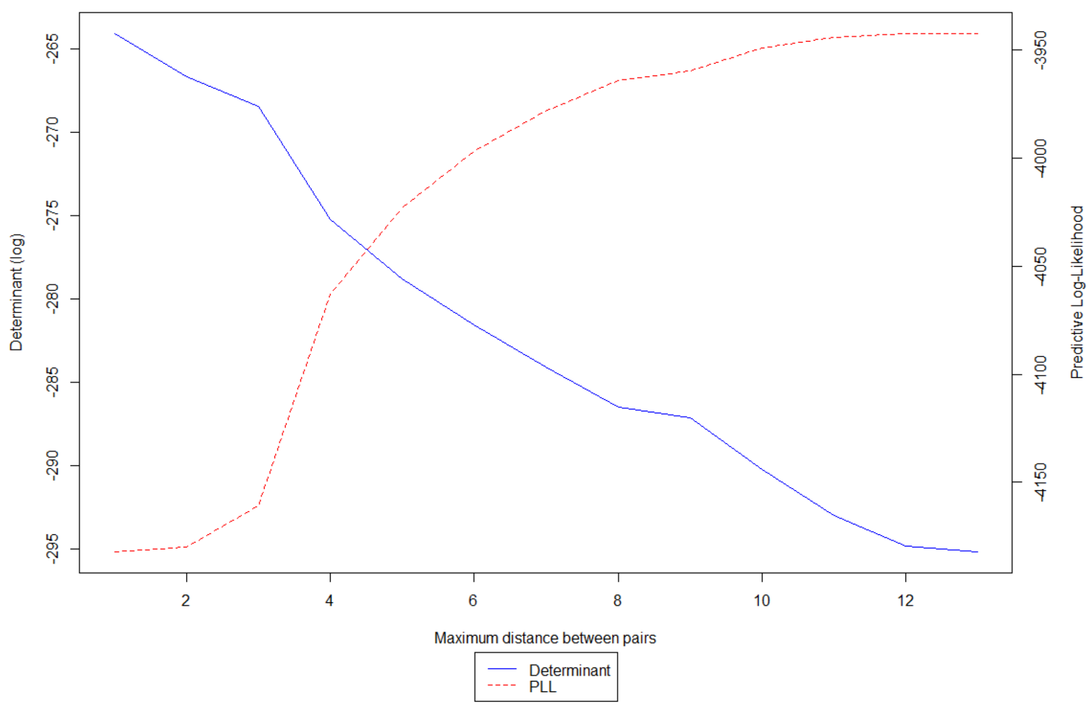

Figure 2: Trace and Predictive Log-Likelihood

In addition of the logarithm of the determinant, we also compute the predictive log-likelihood for each model (Castro et al., 2013), which allows to compare the goodness-of-fit of the CML models to Model A, B, C and D. The results indicate that the adequate distance between pairs is 14 years. Interestingly, the model which features the best value for the logarithm of the determinant is also the model which features the best predictive log-likelihood. Having established the right model composition for Model E, we move on to comparing Model D and Model E.

Table 3: Goodness-of-fit indicators

d	**Determinant**	PLL
2	-264.0736	-4182.26
3	-266.6883	-4179.959
4	-268.4825	-4160.955
5	-275.2321	-4062.872
6	-278.8054	-4022.956
7	-281.5748	-3997.095
8	-284.112	-3978.297
9	-286.5214	-3964.236
10	-287.1156	-3959.526
11	-290.237	-3949.051
12	-292.9751	-3944.075
13	-294.8523	-3942.283
14	-295.1941	-3942.258

The results for Model D and E are reported in Table 4 below. The main result is that the autoregressive specification, Model E, largely outperforms the static one, Model D, in terms of predicting log-likelihood (-6320.049 *versus* -3942.258). In addition, ρ, the parameter which regulates the AR1 process in Model E, is found to be strongly significant. These results clearly indicate that modelling *car availability* as a dynamic process improves model performances. In addition, Model E features smaller standard errors for most parameters in comparison to Model D, which is likely to be induced by the fact that some of the noise in the data is now captured by the autocorrelation parameter, ρ. These quantitative results confirm the qualitative insights of Clark et al. (2016). We now move on to the interpretation of the model parameters, which requires the computation of marginal effects.

5.3 Model interpretation and marginal effects

Parameters interpretation in probit regression is not straightforward. A detailed analysis of the results is further complicated by the generalized structure used in this paper. As previously mentioned, a positive parameter in the *latent car availability propensity* equation means that an increase in the given independent variable leads to an increase in observing a higher outcome. However, a positive parameter in the *Interactions with threshold 2* equation means that an increase in the given independent variable leads to a decrease in observing a higher car availability outcome. To facilitate the interpretation of the results from the *Random effects generalized ordered probit* model and the *AR1 Random effects generalized ordered probit*, we compute the marginal effects[1] for the most

[1] Also known as partial effects in the literature.

Table 4: Model results

Model D					Model E				
Log-Likelihood	-6320.049				Predictive log-likelihood	-3942.258			
AIC BIC	12774.1 13303.78				CM (Log)-Likelihood	-511.5514			
Parameters	*Latent car availability propensity*		*Thresholds and interactions with threshold 2*		Parameters	*Latent car availability propensity*		*Thresholds and interactions with threshold 2*	
	Coefficient	Robust T.	Coefficient	Robust T.		Coefficient	Robust T.	Coefficient	Robust T.
Threshold1	.	.	4.0868	5.2	*Threshold1*	.	.	2.3502	3.47
Threshold2	.	.	-1.1411	-1.6	*Threshold2*	.	.	-1.1609	-2.05
female	-1.754	-4.94	0.151	1.14	*female*	-1.3117	-5.29	0.077	0.57
age	0.0309	2.01	0.0028	0.31	*age*	0.0244	2.52	-0.0044	-0.61
distw	0.0059	0.89	-0.0044	-0.63	*distw*	0.0194	1.95	0.0059	0.91
german	1.032	1.94	0.454	1.34	*german*	0.3021	0.65	0.158	0.5
own_home	0.3183	1.43	0.2536	1.82	*own_home*	0.5352	2.98	0.3407	2.37
moped	-0.8425	-2.29	-0.6148	-1.2	*moped*	-0.3915	-1.18	-0.5741	-1.23
moving	-0.0535	-0.76	-0.0397	-0.89	*moving*	-0.08	-1.59	-0.0702	-1.62
children	0.0722	0.62	0.0296	0.46	*children*	0.0194	0.23	0.077	1.26
birth	0.0946	1.03	0.0667	1.37	*birth*	0.0308	0.49	-0.0284	-0.63
married	0.6628	3.73	0.4809	3.49	*married*	0.4006	2.95	0.3551	2.34
wed	-0.1793	-1.43	-0.1325	-1.42	*wed*	-0.0733	-0.74	-0.0848	-0.96
license_moped	1.2269	2.11	0.0523	0.35	*license_moped*	0.1746	0.75	0.0647	0.47
license_car	3.8801	7.8	0.4813	1.09	*license_car*	2.9127	7.19	0.6356	2.18
central2	-0.1769	-0.61	-0.0034	-0.02	*central2*	0.2009	0.97	0.1167	0.66
central3	-0.3637	-1.29	-0.3096	-1.2	*central3*	0.1561	0.85	-0.1976	-0.98
central4	0.0038	0.01	0.0597	0.29	*central4*	0.2786	1.25	0.2313	1.21
central5	0.3284	1.1	0.0837	0.38	*central5*	0.3571	1.75	0.0948	0.53
pop1	0.2404	0.4	-0.5873	-1.05	*pop1*	-1.0571	-1.87	-1.1706	-1.67
pop2	-0.3853	-0.87	-0.6293	-1.81	*pop2*	-0.4023	-1.22	-0.6954	-2.17
pop3	-0.1441	-0.43	-0.248	-1.28	*pop3*	-0.2372	-0.78	-0.2266	-1.07
pop5	0.2274	0.65	-0.0463	-0.29	*pop5*	-0.0165	-0.04	0.0075	0.03
pop6	0.0806	0.26	-0.1508	-0.93	*pop6*	-0.1118	-0.4	-0.0729	-0.42
pop7	0.1921	0.64	-0.1087	-0.61	*pop7*	-0.2975	-1.01	-0.1589	-0.78
edu_training	1.1061	2.14	-0.1398	-0.8	*edu_training*	0.5777	1.89	-0.0734	-0.4
edu_uni	0.6613	1.32	-0.3864	-1.91	*edu_uni*	0.3598	1.15	-0.3619	-1.7
edu_other	1.5324	2.51	0.0471	0.24	*edu_other*	0.9639	2.62	0.0938	0.45
degree	0.8534	4.13	0.4242	2.4	*degree*	0.4481	2.64	0.3975	2.06
home_det_house	0.2219	1.04	-0.0324	-0.22	*home_det_house*	0.018	0.09	-0.1426	-0.87
home_semi_det_house	0.0244	0.09	-0.1679	-0.87	*home_semi_det_house*	-0.2807	-1.34	-0.3105	-1.56
home_townhouse	0.1154	0.36	-0.0648	-0.36	*home_townhouse*	-0.1619	-0.69	-0.0856	-0.48
home_aprt_big	0.116	0.29	0.189	0.56	*home_aprt_big*	0.0011	0	0.1465	0.66
home_other	-0.5194	-2.15	-0.6047	-2.23	*home_other*	-0.3868	-1.75	-0.7562	-2.78
σ *(rand. effect)*	2.9987	14.7	.	.	σ *(rand. effect)*	1.623	7.04	.	.
ρ	.	.	.	.	*Rho*	2.6441	11.54	.	.

significant parameters. Marginal effects can simply be described as the effect of a discrete variation of a given predictor x on the probability of observing the modelled outcome all else being equal.

5.3.1 Computation

Details on the computation of marginal effects in the context of probit models are given by Greene and Hensher (2010). Following the notations established in Equation 4 and accounting for the fact that most of the variables used in this analysis appear in both $Y*$ and μ_2, it comes that

$$\frac{\delta Prob(Y=k|x,z)}{\delta x} = (\phi(\mu_k - \beta' x) - \phi(\mu_{k-1} - \beta' x))\beta \tag{13}$$

and

$$\frac{\delta Prob(Y=k|x,z)}{\delta z} = (\phi(\mu_k - \beta' x)\mu k \zeta_k - \phi(\mu_{k-1} - \beta' x))\mu k - 1\zeta_{k-1} \tag{14}$$

We compute the marginal effects for the three outcome considered ([Car] *"0. Never available"*, *"1. Sometimes available"* and *"2. Always available"*) and the models with and without autocorrelation. This allows us to give richer comments on the differences in terms of behavioural interpretations provided by the two models as well as to further demonstrate the need to account for autocorrelation in large panel datasets such as those typically found in life-course analysis. Results are reported in Table 5 below. Given that the marginal effects are computed "at means", that is by assuming during computations that the values for all the predictors is at their corresponding mean, some marginal effects are not found to be significant although their corresponding parameters in Table 4 are. In this paper, we only compute the marginal effects for the parameters which have been found to be significant in the model estimation phase.

5.3.2 Interpretation

We describe how to interpret the results from Table 5 by taking the example of the variable *female* for Model E. According to the model results, being a woman with respect to being a man increases the probability of reporting a car as *never available* by 0.1972 while it decreases the probability of reporting a car as *Always available* by 0.3135. For continuous variables (*age* and *distw* (distance to work)), the values correspond to a shift in outcome probability for a unit increase.

We report differences between Model D and Model E. In particular, we find that the marginal effects derived from both models are similar in sign, but that the simple random effects models fails to capture effects which are typically considered to be strongly significant in the literature in some cases. For example, the variable $female$, for which

Table 5: Marginal effects (at means)

Parameters	Outcome	Random effects generalized order probit			AR1 random effects generalised ordered probit		
		Marginal effect	Std. Dev.	T-ratio	Marginal effect	Std. Dev.	T-ratio
female	*0. Never*	0.1923	0.1479	1.3006	0.1972	0.1002	1.9687
	1. Sometimes	0.0647	0.1013	0.6390	0.1163	0.0777	1.4963
	2. Always	-0.2570	0.1150	-2.2344	-0.3135	0.0956	-3.2794
age	*0. Never*	-0.0035	0.0028	-1.2707	-0.0039	0.0017	-2.2972
	1. Sometimes	-0.0002	0.0024	-0.0979	-0.0029	0.0025	-1.1721
	2. Always	0.0037	0.0027	1.3829	0.0068	0.0028	2.4516
distw	*0. Never*	-0.0007	0.0009	-0.7084	-0.0031	0.0021	-1.4586
	1. Sometimes	-0.0007	0.0009	-0.7604	-0.0001	0.0015	-0.0661
	2. Always	0.0014	0.0008	1.7061	0.0032	0.0014	2.3389
own_home	*0. Never*	-0.0360	0.0329	-1.0933	-0.0842	0.0468	-1.8018
	1. Sometimes	0.0449	0.0709	0.6343	0.0435	0.0508	0.8568
	2. Always	-0.0089	0.0779	-0.1147	0.0408	0.0576	0.7070
married	*0. Never*	-0.0776	0.0760	-1.0213	-0.0679	0.0423	-1.6034
	1. Sometimes	0.0741	0.0834	0.8885	0.0530	0.0467	1.1349
	2. Always	0.0035	0.0921	0.0384	0.0149	0.0527	0.2826
licence_car	*0. Never*	-0.4820	0.0383	-12.5819	-0.6202	0.0855	-7.2545
	1. Sometimes	0.0932	0.1805	0.5161	0.1040	0.1289	0.8071
	2. Always	0.3888	0.1748	2.2243	0.5162	0.1031	5.0073
central5	*0. Never*	-0.0360	0.0346	-1.0415	-0.0510	0.0335	-1.5231
	1. Sometimes	0.0099	0.0900	0.1104	-0.0071	0.0512	-0.1394
	2. Always	0.0261	0.0901	0.2899	0.0581	0.0591	0.9839
pop1	*0. Never*	-0.0266	0.0447	-0.5958	0.2138	0.1961	1.0900
	1. Sometimes	-0.0910	0.1774	-0.5128	-0.1330	0.1225	-1.0858
	2. Always	0.1176	0.2857	0.4118	-0.0808	0.2274	-0.3551
pop2	*0. Never*	0.0452	0.0779	0.5803	0.0713	0.0644	1.1058
	1. Sometimes	-0.0870	0.1212	-0.7182	-0.0998	0.0719	-1.3879
	2. Always	0.0418	0.1531	0.2732	0.0286	0.0991	0.2881
edu_training	*0. Never*	-0.1108	0.0528	-2.0989	-0.0767	0.0449	-1.7101
	1. Sometimes	-0.0545	0.1461	-0.3730	-0.0647	0.0741	-0.8740
	2. Always	0.1653	0.1748	0.9462	0.1415	0.0953	1.4848
edu_university	*0. Never*	-0.0700	0.0495	-1.4138	-0.0513	0.0413	-1.2442
	1. Sometimes	-0.0764	0.1494	-0.5116	-0.0952	0.0826	-1.1529
	2. Always	0.1464	0.1765	0.8293	0.1465	0.1029	1.4244
edu_other	*0. Never*	-0.1451	0.0540	-2.6860	-0.1121	0.0514	-2.1802
	1. Sometimes	-0.0386	0.1463	-0.2636	-0.0681	0.0792	-0.8602
	2. Always	0.1836	0.1757	1.0449	0.1802	0.0998	1.8049
degree	*0. Never*	-0.0883	0.0468	-1.8848	-0.0621	0.0355	-1.7505
	1. Sometimes	0.0804	0.1025	0.7841	0.0771	0.0713	1.0804
	2. Always	0.0078	0.1152	0.0681	-0.0149	0.0780	-0.1916
home_other	*0. Never*	0.0616	0.0771	0.7980	0.0683	0.0555	1.2293
	1. Sometimes	-0.0831	0.1092	-0.7614	-0.1074	0.0707	-1.5189
	2. Always	0.0216	0.1239	0.1740	0.0391	0.0864	0.4531

the marginal effect related to the outcome *never available* is significant in the AR1 case but not in the simple random effect case. This result is also found for *age* as well as *own_home*. This result is concerning in the sense that it shows that not accounting for autocorrelation in life-course analysis can lead to misguided interpretations on the marginal effect of certain variables as crucial as gender and age. In what follows, we focus on interpreting the results from the AR1 model.

As previously mentioned, we find a very strong gender effect. Women are much less likely that men to report a car as *always available* and much more likely to report a car as *never available*. This is a common result in the literature (Cao et al., 2007; Scheiner and Holz-Rau, 2012; Simma and Axhausen, 2001). Scheiner and Holz-Rau (2012) suggest that social roles and different access to resources inside and outside the household can explain this result, among other factors. In addition, age is found to negatively influence the probability of reporting a car as *never available* and increases the probability of reporting a car as *always available*. This is a result which is in line with Dargay (2002) and Prillwitz et al. (2006), among others. More precisely, Dargay (2002) enunciates that car ownership grows rapidly as the age of the household head increases, reaches a peak at around 50 and slowly decreases thereafter. A similar non-linear relationship is found by Prillwitz et al. (2006). There are two reasons for us to not observe such a pattern. Firstly, our sample is not representative and, as a result, might not feature as many older respondents than other studies. Moreover, these studies investigate car ownership while we investigate car availability, which means that even if the number of cars owned by respondents decreases passed a certain threshold, this might not have an impact of the availability of the remaining car(s). Finally, the drop in car availability observed in Figure 1 after 2010 is misleading in the sense that it is due to sample attrition and not necessarily to a tendency of the respondents to experience a reduced car availability level as they get older.

Home owners are less likely to report a car as *never available* (-0.0842). This might be driven by the fact that home ownership is connected with higher financial resources, and that according to Clark et al. (2016), *"affective desire for cars may arise from a change in resources e.g. increased income prompting a greater desire for a 'better' car"*, or, in the context of our study, an increase in car availability status. Being married is found to have a strong and significant positive effect on latent car availability propensity, but this doesn't lead to significant marginal effects. This might be driven by the fact that the effect of *married* on μ_2 is also positive, which means that being married decreases the chances of reporting a car as *always available* all else being equal, making the overall effect of being married complex to capture. Having a car driving licence largely decreases the probability of reporting a car as *never available* (-0.62) and increases the probability of reporting a car as *always available*, which is also in line with the literature (Whelan, 2007).

Geographical factors are also found to affect the latent car availability propensity, which is an extremely common result in the literature (Guerra, 2015). Living in a periurban area (*central*5) has a positive effect on the latent outcome while living in a small city (*pop*1 and *pop*2) decreases the value of μ_2 but also decreases the value of the latent outcome, which complicates again the interpretation of these effects. We note that the marginal effects are not found to be significant for these variables.

The effect of education level on latent car availability is found to be positive with respect to the base, which is high school education only. Given that education can be used as a proxy for income, this is again in line with previous findings (Clark et al., 2016). The effect of having a university degree is ambiguous given that *degree* has a positive effect on latent car availability but also decreases the chances of reporting a car as *always available*.The marginal effects indicate that having a degree reduces the probability of reporting a car as *never available* by 0.0683.

Finally, the marginal effects are interpreted by the means of graphical representations. Figure 3 reports the marginal effects for *age* and *distw*, which are continuous, while 4 addresses shifts in discrete outcomes. We find that the effect of *age* and *distw* becomes particularly important as the variables increase. As an illustration, if *age* increases by 25 all else being equal, the probability of reporting a car as *always available* increases by 0.2.

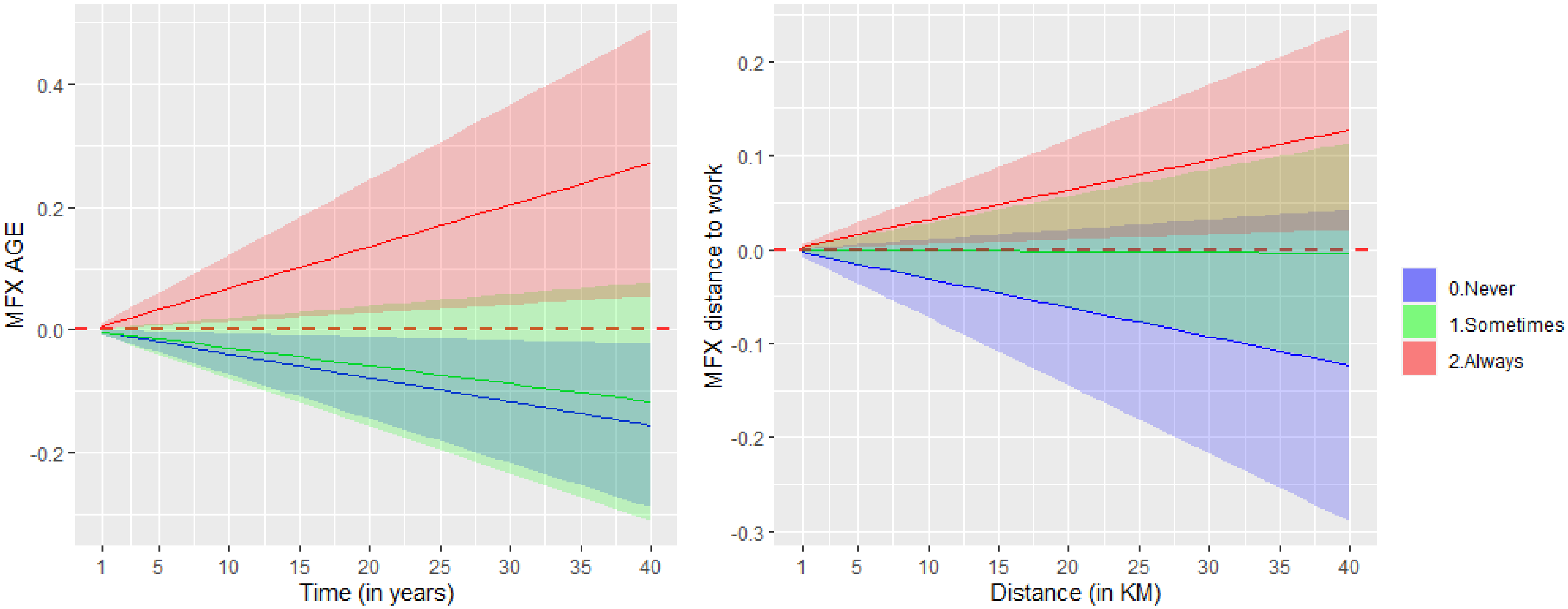

Figure 3: Marginal effects - Continuous variables

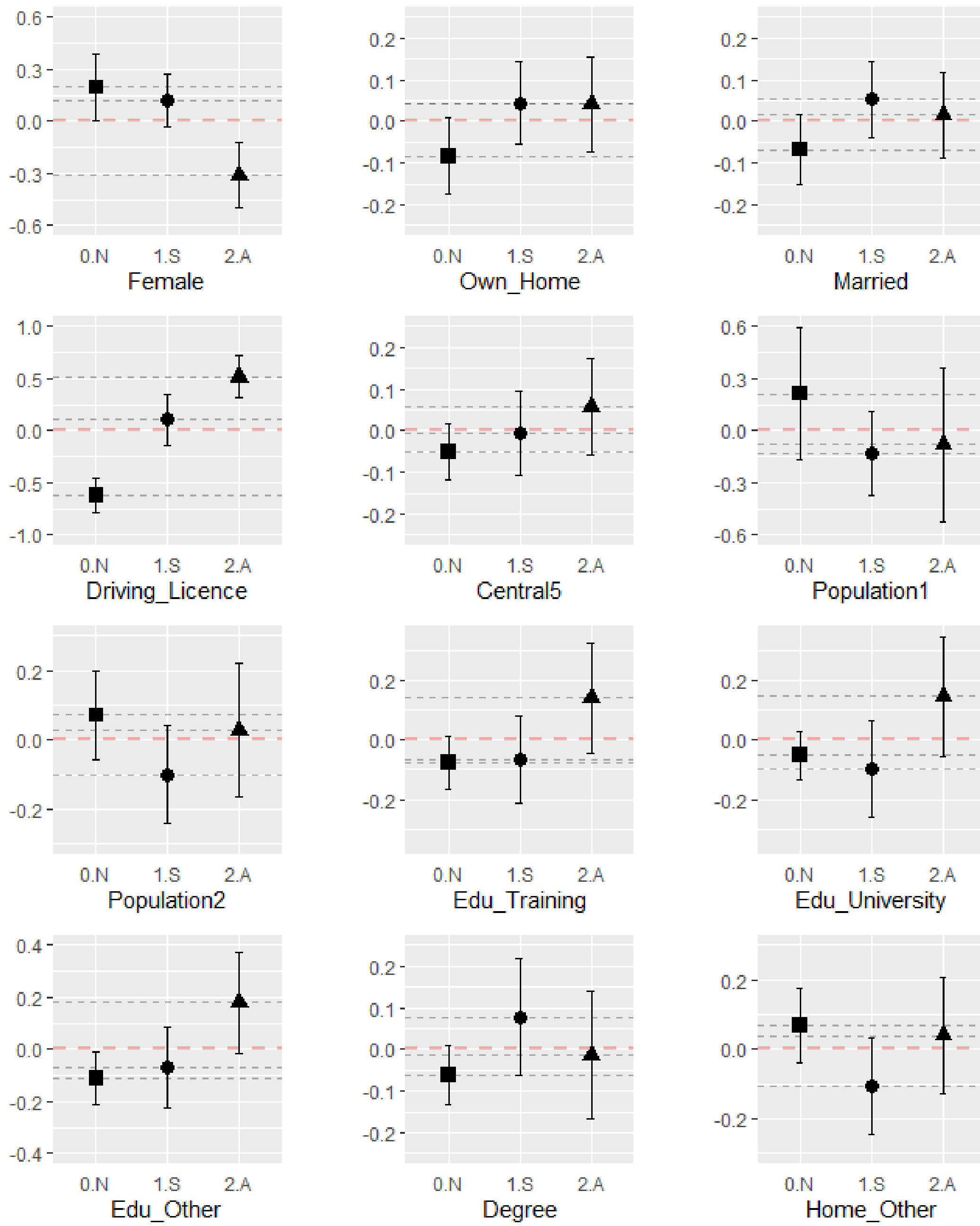

Figure 4: Marginal effects - Discrete variables

6 Conclusions and Outlook

This paper analysed the determinants of household car availability in Germany using data from a life course calendar survey that took place in Dortmund between 2007 and 2012. Car availability is a common focus in the life course calendar literature and understanding the dynamics of car availability and, on a broader context, of mobility and mobility tool choice is crucial for policy design. In contrary to similar approaches on the same topic, our data cover an extensive period of time because of the use of a life course calendar approach for collecting data (up to 51 years per individual). A particular focus of the paper was to compare the modelling results that are obtained following common practices in the car availability literature, based on unordered response structures, with the results obtained with more recent econometric approaches such as an autoregressive generalised ordered choice model, which as we demonstrate, assumes a process which is more in line with recent qualitative findings (Clark et al., 2016).

In this paper we have introduced the autoregressive model to the examination of the life course, and the initial results are such that this approach shows great promise as a method. In particular, we first suggest to extend the use of the model to investigate a wider range of choices that are of interest in the life course calendar literature.

The autoregressive approach may be seen as a superior alternative in the context of analysing car availability and, in a broader context, life course events for the main reason that it accounts for state dependency. Hence, adopting a dynamic approach consists in asking whether car availability status in past periods affects present car availability. We argue that state dependency is a very important aspect to consider in the context of life course calendar analysis in the sense that most of the past long-term decisions made by individuals such as buying a house or changing job affect their preferences in future periods and induce economic constraints in the form of transaction costs.

7 Acknowledgements

This research was funded by DFG / SNF as part of the project "Mobility Biographies: A Life-Course Approach to Travel Behaviour and Residential Choice" - a collaborative of TU Dortmund, Goethe University Frankfurt and ETH Zurich. Additionally, the authors acknowledge the financial support by the European Research Council through the consolidator grant 615596-DECISIONS.

References

Axhausen, K. W. (2008). Social networks, mobility biographies, and travel: survey challenges. *Environment and Planning B: Planning and design*, 35(6):981–996.

Beige, S. and Axhausen, K. W. (2008). Long-term and mid-term mobility decisions during the life course: experiences with a retrospective survey. *IATSS research*, 32(2):16–33.

Beige, S. and Axhausen, K. W. (2012). Interdependencies between turning points in life and long-term mobility decisions. *Transportation*, 39(4):857–872.

Bhat, C. R. (2011). The maximum approximate composite marginal likelihood (macml) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B: Methodological*, 45(7):923–939.

Bhat, C. R. et al. (2014). The composite marginal likelihood (cml) inference approach with applications to discrete and mixed dependent variable models. *Foundations and Trends® in Econometrics*, 7(1):1–117.

Bhat, C. R. and Pulugurta, V. (1998). A comparison of two alternative behavioral choice mechanisms for household auto ownership decisions. *Transportation Research Part B: Methodological*, 32(1):61–75.

Cao, X., Mokhtarian, P. L., and Handy, S. L. (2007). Cross-sectional and quasi-panel explorations of the connection between the built environment and auto ownership. *Environment and Planning A*, 39(4):830–847.

Castro, M., Paleti, R., and Bhat, C. R. (2013). A spatial generalized ordered response model to examine highway crash injury severity. *Accident Analysis & Prevention*, 52:188–203.

Clark, B., Lyons, G., and Chatterjee, K. (2016). Understanding the process that gives rise to household car ownership level changes. *Journal of Transport Geography*, 55:110–120.

Dargay, J. M. (2001). The effect of income on car ownership: evidence of asymmetry. *Transportation Research Part A: Policy and Practice*, 35(9):807–821.

Dargay, J. M. (2002). Determinants of car ownership in rural and urban areas: a pseudo-panel analysis. *Transportation Research Part E: Logistics and Transportation Review*, 38(5):351–366.

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*, pages 287–301.

Döring, L., Kroesen, M., and Holz-Rau, C. (2019). The role of parents' mobility behavior for dynamics in car availability and commute mode use. *Transportation*, 46(3):957–994.

Eluru, N. (2013). Evaluating alternate discrete choice frameworks for modeling ordinal discrete variables. *Accident Analysis & Prevention*, 55:1–11.

Eluru, N., Bhat, C. R., and Hensher, D. A. (2008). A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accident Analysis & Prevention*, 40(3):1033–1054.

Erickson, B. H. (1979). Some problems of inference from chain data. *Sociological Methodology*, 10(1):276–302.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, 31(4):1208–1211.

Golob, T. F. (1990). The dynamics of household travel time expenditures and car ownership decisions. *Transportation Research Part A: General*, 24(6):443–463.

Golob, T. F. and Van Wissen, L. (1989). A joint household travel distance generation and car ownership model.

Greene, W. H. and Hensher, D. A. (2010). *Modeling ordered choices: A primer*. Cambridge University Press.

Guerra, E. (2015). The geography of car ownership in mexico city: a joint model of households' residential location and car ownership decisions. *Journal of Transport Geography*, 43:171–180.

Hanly, M. and Dargay, J. M. (2000). Car ownership in great britain: Panel data analysis. *Transportation Research Record*, 1718(1):83–89.

Hensher, D. A. (2013). *Dimensions of automobile demand: a longitudinal study of household automobile ownership and use*. Elsevier.

Hensher, D. A. and Le Plastrier, V. (1985). Towards a dynamic discrete-choice model of household automobile fleet size and composition. *Transportation Research Part B: Methodological*, 19(6):481–495.

Ierza, J. V. (1985). Ordinal probit: a generalization. *Communications in Statistics-Theory and Methods*, 14(1):1–11.

Joe, H. and Lee, Y. (2009). On weighting of bivariate margins in pairwise likelihood. *Journal of Multivariate Analysis*, 100(4):670–685.

Jong, G. D., Fox, J., Daly, A., Pieters, M., and Smit, R. (2004). Comparison of car ownership models. *Transport Reviews*, 24(4):379–408.

Kitamura, R. (1987). A panel analysis of household car ownership and mobility. *Doboku Gakkai Ronbunshu*, 1987(383):13–27.

Kitamura, R. and Bunch, D. S. (1990). Heterogeneity and state dependence in household car ownership: A panel analysis using ordered-response probit models with error components.

Klöpper, V. and Weber, A. (2007). Generationsübergreifende mobilitätsbiographien. Master's thesis, Faculty of Spatial Planning, Technische Universität Dortmund.

Lanzendorf, M. (2003). Mobility biographies: A new perspective for understanding travel behaviour. In *Paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, August 2003*.

Maddala, G. S. (1986). *Limited-dependent and qualitative variables in econometrics*. Number 3. Cambridge university press.

Mannering, F. L. (1983). An econometric analysis of vehicle use in multivehicle households. *Transportation Research Part A: General*, 17(3):183–189.

Manzoni, A., Vermunt, J. K., Luijkx, R., and Muffels, R. (2010). 2. memory bias in retrospectively collected employment careers: A model-based approach to correct for measurement error. *Sociological methodology*, 40(1):39–73.

Meurs, H. (1993). A panel data switching regression model of mobility and car ownership. *Transportation Research Part A: Policy and Practice*, 27(6):461–476.

Müggenburg, H., Busch-Geertsema, A., and Lanzendorf, M. (2015). Mobility biographies: A review of achievements and challenges of the mobility biographies approach and a framework for further research. *Journal of Transport Geography*, 46:151–163.

Nobile, A., Bhat, C. R., and Pas, E. I. (1997). A random-effects multinomial probit model of car ownership choice. In *Case studies in Bayesian statistics*, pages 419–434. Springer.

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.

Paleti, R. and Bhat, C. R. (2013). The composite marginal likelihood (cml) estimation of panel ordered-response models. *Journal of choice modelling*, 7:24–43.

Pauli, F., Racugno, W., and Ventura, L. (2011). Bayesian composite marginal likelihoods. *Statistica Sinica*, pages 149–164.

Prillwitz, J., Harms, S., and Lanzendorf, M. (2006). Impact of life-course events on car ownership. *Transportation Research Record*, 1985(1):71–77.

Scheiner, J. (2007). Mobility biographies: Elements of a biographical theory of travel demand (mobilitätsbiographien: Bausteine zu einer biographischen theorie der verkehrsnachfrage). *Erdkunde*, pages 161–173.

Scheiner, J. (2014). Gendered key events in the life course: effects on changes in travel mode choice over time. *Journal of Transport Geography*, 37:47–60.

Scheiner, J. and Holz-Rau, C. (2012). Gender structures in car availability in car deficient households. *Research in Transportation Economics*, 34(1):16–26.

Scheiner, J. and Holz-Rau, C. (2013a). Changes in travel mode use after residential relocation: a contribution to mobility biographies. *Transportation*, 40(2):431–458.

Scheiner, J. and Holz-Rau, C. (2013b). 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.

Scheiner, J., Sicks, K., and Holz-Rau, C. (2014). Generationsübergreifende mobilitätsbiografien-dokumentation der datengrundlage: Eine befragung unter studierenden, ihren eltern und großeltern (arbeitspapiere des fachgebiets verkehrswesen und verkehrsplanung 29). *working paper of the Faculty of Spatial Planning, TU Dortmund*, 29.

Simma, A. and Axhausen, K. W. (2001). Structures of commitment in mode use: a comparison of switzerland, germany and great britain. *Transport Policy*, 8(4):279–288.

Srinivasan, K. K. (2002). Injury severity analysis with variable and correlated thresholds: ordered mixed logit formulation. *Transportation Research Record*, 1784(1):132–141.

Varin, C. (2008). On composite marginal likelihoods. *AStA Advances in Statistical Analysis*, 92(1):1–28.

Varin, C. and Czado, C. (2009). A mixed autoregressive probit model for ordinal longitudinal data. *Biostatistics*, 11(1):127–138.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, pages 5–42.

Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, 92(3):519–528.

Whelan, G. (2007). Modelling car ownership in great britain. *Transportation Research Part A: Policy and Practice*, 41(3):205–219.

Zhang, J., Yu, B., and Chikaraishi, M. (2014). Interdependences between household residential and car ownership behavior: a life history analysis. *Journal of Transport Geography*, 34:165–174.

Zhao, Y. and Joe, H. (2005). Composite likelihood estimation in multivariate data analysis. *Canadian Journal of Statistics*, 33(3):335–356.