Weighting strategies for modelling life course history events *via* pairwise composite marginal likelihood

Romain Crastes dit Sourd[*†] David Palma[*‡] Stephane Hess[*§]
Andrew Daly[*¶] Christian Holz-Rau[‖**] Joachim Scheiner[‖††]

August 26, 2020

WORKING PAPER - DO NOT QUOTE WITHOUT PERMISSION.

Abstract

Models using large temporal datasets often feature complex likelihood functions. In the context of life course history data, the dimension of integration may increase with the number of time periods considered which rules out Maximum Simulated Likelihood (MSL) estimation techniques. An alternative is to use the Composite Marginal Likelihood (CML) inference approach, which replaces high-dimensional integrals by a compounding of bivariate probabilities. The current paper delves into the issue of how to form CML functions. CML is a flexible tool which allows to use different weights on different bivariate margins (typically 0 or 1) which will affect goodness-of-fit, efficiency as well as computation speed. The typical approach consists of using the pairing combinations of temporally close choice situations and assign a weight of 0 to the other pairs. Other weighting strategies such as using non-rectangular or random weights are possible but rarely used In this paper, we propose a large simulation exercise based on a real dataset on car availability over the life course in Germany to compare the finite-sample performances of the CML approach under alternative weighting strategies. Our results indicate that the typical weighting approach is outperformed in a large majority of cases. We also find that introducing observed and random heterogeneity in weights improves model performances in terms of parameter recovery. Finally, we unravel the important role played by spurious state dependence in car availability across the life course.

Keywords: car availability, mobility biographies, composite marginal likelihood estimation

[*]Institute for Transport Studies and Choice Modelling Centre, University of Leeds (UK)
[†]r.crastesditsourd@leeds.ac.uk
[‡]d.palma@leeds.ac.uk
[§]s.hess@leeds.ac.uk
[¶]andrew@alogit.com
[‖]TU Dortmund, Germany
[**]christian.holz-rau@tu-dortmund.de
[††]joachim.scheiner@tu-dortmund.de

1 Introduction

It is well known that dependence on past travel habits and life course experiences affects current travel behaviour (Hanly and Dargay, 2000). It is essential to identify the most influential factors of travel behaviour over time in order to contribute to planning practice (Axhausen, 2008). In the past decade, the focus of interest has therefore shifted towards individual decisions in a life span. In a recent comprehensive review of the theoretical framework and most important studies investigating mobility behaviour and mobility tool ownership over the life course, Müggenburg et al. (2015) highlight that studies often investigate mobility decisions with static (panel) models and should not neglect the temporal dimension of decision-making because it may lead to biased estimates, in particular with an inappropriate treatment of state dependence. State dependence can be "true" or "spurious" (Heckman, 1981). True state dependence simply describes the fact that past observed mobility choices influence future decisions. Spurious state dependence refers to the correlation between model errors across time periods (autocorrelation), caused by the persistence of tastes.

Modelling state dependence in the context of car availability, which is the empirical context of the current study, or car ownership has received much attention in the past (see Cherchi and Manca (2011); Dargay and Hanly (2007); Golob (1990); Hensher (2013); Kitamura and Bunch (1990); Manski and Sherman (1980) and Clark et al. (2009) among others). However, while there exists an abundant literature on true state dependence and how to account for it, spurious state dependence is rarely considered in panel choice models, which may be due to the fact that the estimation of such models is usually particularly burdensome. The basic approach for modelling repeated choices in the literature is to introduce random effects which generate correlations across time for the same individual. This does not only apply to the case where the dependent variable is a binary outcome but also to ordered dependent variables or count data. Random effects models are generally estimated using Maximum Simulated Likelihood (MSL) to approximate the multi-variate integrals contained in the likelihood function (Hensher and Greene, 2003; Paleti and Bhat, 2013; Train, 2001). Numerical simulation methods can be very time consuming and may even not reach a solution, especially when modelling spurious state dependence, where accounting for autocorrelation requires an analyst to evaluate multiple integrals (one per time period considered). 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).

CML is an estimation technique which has been developed, refined and used in a wide range of contexts. More precisely, CML has been extensively used to model spatial interaction effects. As an illustration, Bhat et al. (2010) proposed to use CML in order to account for spatial dependence in a model of teenagers' weekday recreational activity participation. Narayanamoorthy et al. (2013) proposed a multivariate model for the analysis of injury counts by road-user type and injury severity level which accounts for spatial dependence effects in injury counts. CML has also been used to jointly model different types of choices and dependent variables. For example, Paleti et al. (2013) presented an integrated model of residential location, work location, vehicle ownership and commute tour characteristics. More recently, Bhat (2015) jointly modelled multiple nominal outcomes, multiple ordinal variables, multiple count variables and multiple continuous variable in order to integrate residential location choices and travel behaviour. A generalized heterogeneous data model of mixed types of dependent variables which can also accommodate spatial interactions has also been proposed by Bhat et al. (2016). CML can also accommodate both spatial and temporal dependences within the same modelling framework (Castro et al., 2012), and has been successfully applied to the modelling of multiple discrete-discrete and multiple discrete-continuous choices (Bhat et al., 2013; Enam et al., 2018). Finally, it is worth noting that the use of the multivariate skew-normal distribution allows to accommodate non-normal mixing within a CML framework (Bhat and Sidharthan, 2012).

A typical CML function is the full-pairwise likelihood, which gives the same weight w (generally 1) to each pair of observed choice situation from the same individual. This approach requires only the evaluation of a series of bivariate normal probabilities rather than a single high-dimensional integral. However, the literature on CML suggests that it may not be necessary to evaluate all the possible pairs but only the closest ones, and assign a weight of 0 to other pairs. More precisely, Varin and Czado (2009), among others, show that considering pairs which are too distant (*i.e.* more than a given number of time units apart) may not only increase the computational burden but also reduce the model fit. The optimal distance between pairs is usually found by estimating a series of models using increasingly distant pairs and computing the trace (sum of diagonal elements) of the asymptotic variance covariance matrix for each model, with the distance which minimises the trace selected as the *optimal* distance. We label this approach the distance-based rectangular weighting approach and we note that it appears to be the leading approach in the literature (Bhat et al., 2010; Castro et al., 2012; Ferdous and Bhat, 2013; Sang and Genton, 2014; Varin and Czado, 2009). It provides obvious speed gains and convergence improvements but does not appear to improve efficiency further. A recent contribution from Papageorgiou and Moustaki (2018) proposes a new sampling strategy which improves computation time but does not affect efficiency. More recently, Bevilacqua et al. (2012) have reported that non-rectangular weights (which are bounded between 1 and 0 and allow to gradually decrease the contribution to the likelihood as distance between pairs increases) improve the efficiency of maximum pairwise likelihood estimators. In addition, Pedeli and Varin (2018) have also suggested to use non-rectangular weights based on kernel functions although they could not report any gain in efficiency.

In their discussion, Paleti and Bhat (2013) indicate that how exactly to form a CML function remains a wide open area, and this forms a key motivation for the present paper. Using a unique mobility biography data-set collected in Dortmund in 2007, we analyse the determinants of car availability for 437 German women between 1980 and 2006 by the means of an autoregressive random-effects ordered probit model. The CML formulation of this model has been initially proposed by (Varin and Czado, 2009) and (Paleti and Bhat, 2013). We explore the suggestion made by Paleti and Bhat (2013) among others and compare the performances of the CML estimator under a large different set of weighting strategies including distance-based rectangular weights, non-rectangular weights based on decay functions as well as random and semi-random weights. More precisely, we estimate a series of CML models using the traditional distance-based rectangular weighting approach and find the optimal distance between pairs. Then, we generate a series of simulated datasets using the results from the model using the optimal distance between pairs and the real dataset. We use these simulated datasets to compare the finite-sample performance of the CML approach under different weighting strategies. Our results reveal that the traditional distance-based rectangular weighting approach is outperformed by a large majority of alternatives weighting strategies, and that introducing heterogeneity (observed and random) in the weighting process yields better results in terms of parameter recovery.

The remainder of this paper is structured as follows. The next section presents the MSL estimation method and the CML estimation method of the autoregressive random-effects ordered probit model and discusses different methods for selecting models estimated *via* CML. This is followed by an overview of the data used in our case study and details on model specification. Finally, we discuss the results and present some conclusions.

2 Autoregressive ordered probit models

In this section we describe the equations underlying the MSL and CML estimation of the autoregressive random effects ordered probit model. We build our models step-by-step by first introducing the existing MSL and CML formulations of the random-effects ordered probit model (*reoprobit*). We then present the autoregressive versions of this model (Paleti and Bhat, 2013). In addition, we provide some existing and new insights on model composition and selection for CML estimation.

2.1 The maximum simulated likelihood random effects ordered probit model

The *reoprobit* is a simple extension of the ordered probit model (*oprobit*) which accommodates panel data. In this paper, we provide only a short description of this well-known model in order to better introduce the more complex autoregressive models we use in this paper. The model can be specified as follows:

$$y^*_{ij} = \beta' X_{ij} + \alpha_i + \epsilon_{ij}, \tag{1}$$

where y^*_{ij} is the unobserved latent outcome defined as a function of relevant exogenous variables, i is an index for individuals (i = 1, 2, ..., I) and j is an index for the j^{th} observation with j = 1, 2, ..., J the number of periods under study, X_{ij} is a vector of exogenous variables, and β' is a vector of coefficients to be estimated while 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[1]. We have 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 $\mu_K = +\infty$. With this notation, μ_1, μ_2, ..., μ_{K-1} are parameters to be estimated with $\mu_1 < \mu_2 < ... < \mu_{K-1}$. 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 *reoprobit* model is easily and rapidly estimated using 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:

$$P(k_i)|\eta_i = \prod_{j=1}^{J} \left(\Phi(\mu_{ijk} - \alpha - \beta' X_{ijk-1} - \eta_i) - \Phi(\mu_{ij} - \alpha - \beta' X_{ij} - \eta_i) \right) \tag{2}$$

[1] We acknowledge the fact that the use of an ordered probit model for this choice is a bit contentious as it assumes that the key effects separating never available from sometimes available are the same as the effects separating sometimes available from always available. The use of an ordered probit model has been motivated by the methodological scope of this paper. Future work on this topic will explore the use of alternative models such as the multinomial probit.

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.

$$logL_i(\theta) = \\ log\Bigg[\int_{-\infty}^{+\infty}\prod_{j=1}^{J}\bigg(\Phi(\mu_{ijk}-\alpha-\beta' X_{ij}-\sigma v)-\Phi(\mu_{ijk-1}-\alpha-\beta' X_{ij}-\sigma v)\bigg)\phi(v)dv\Bigg] \tag{3}$$

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 *reoprobit* model entails only a one dimensional integral so model estimation is generally fast. Moreover, the model is also not prone to convergence related issues and there are thus no known reasons for estimating the *reoprobit* model using CML.

2.2 The maximum simulated likelihood autoregressive random effects ordered probit model

The simple *reoprobit* model presented above assumes that the multiple observations for each individual are equally correlated across time. However, in many cases, the correlation across errors varies across time. This is particularly relevant in the context of car availability where the errors are likely to be more correlated for closer observations. The *areoprobit* can accommodate an autoregressive structure of order one for the error term (for each individual) but this leads to substantial increases in computational efforts when estimated using MSL. 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 (3) only entails a one-dimension integral, the autoregressive model requires the evaluation of an integral of dimension J for individual i. The log-likelihood function becomes:

$$logL_i(\theta) = \bigg[\int_{w_1=\delta_{m_{i1}-1}}^{\delta_{m_{i1}}}, ..., \int_{w_J=\delta^{m_{iJ}-1}}^{\delta_{m_{iJ}}} \phi_J(w_1, ..., w_J|\Sigma)dw_1, ..., dw_J\bigg], \tag{4}$$

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 *areoprobit*. For example, in the context of the application presented in this paper, the full information likelihood estimation has the order of 27 dimensions of integrations. Such a model would take weeks to converge and would be very prone to simulation errors (Paleti and Bhat, 2013). These issues are easily circumvented by the CML estimation approach, which, in the context of this paper, entails only the evaluation of pairs of bivariate normal probabilities.

2.3 The composite marginal likelihood autoregressive random effects ordered probit model

As previously mentioned, 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:

$$logL_i(\theta) = \sum_{g=j+1}^{J} \sum_{j=1}^{J-1} log\Big[Pr(y_{ij} = k_{ij}, y_{ig} = k_{ig})\Big], \quad (5)$$

where

$$\begin{aligned} &Pr(y_{ij} = k_{ij}, y_{ig} = k_{ig}) \\ &= \phi_2(\delta_{k_{ij}}, \delta_{k_{ig}}, \zeta_{ig}) - \phi_2(\delta_{k_{ij}}, \delta_{k_{ig-1}}, \zeta_{ig}) \\ &- \phi_2(\delta_{k_{ij-1}}, \delta_{k_{ig}}, \zeta_{ig}) + \phi_2(\delta_{k_{ij-1}}, \delta_{k_{ig-1}}, \zeta_{ig}) \end{aligned} \quad (6)$$

It is worth noting that (6) 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 7 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. Different pair selection strategies exist in the literature and the full-pairwise approach is only one of them.

3 Weighting strategies

In this section, we discuss the choice of pairs used in CML estimation. We first look at the approach which is the most widely used in the literature before turning our head to existing and new alternatives.

3.1 Constant and distance-based rectangular weights

As previously discussed in the introduction, the full-pairwise marginal likelihood function also presented in equation (7) 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. However, the full-pairwise approach becomes more computationally intensive as the number of time periods increases. A full-pairwise approach applied to a hypothetical balanced dataset featuring 10 time periods per individual requires the evaluation of 45 pairs of bivariate normal probabilities. This number increases to 190 when 20 time periods need to be considered and 1,225 for a 50 time periods case, quickly becoming intractable. 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

$$logL_i, d(\theta) = \sum_{g=j+1}^{J} \sum_{j=1}^{d} w_{ijg} logPr(y_{ij} = k_{ij}, y_{ig} = k_{ig}) \qquad (7)$$

In the large majority of the literature (REF), the weights w_{ijg} (which corresponds to the weights associated to the pair composed of thevent observed at time j and g for individual i) are of type

$$w_i ijg = \begin{cases} 1, 1 \leq t \leq d, \\ 0, \text{otherwise} \end{cases}$$

However, different choices of the weights might also be possible, some of which have been already used in the literature.

3.2 Non-rectangular weights

Very few studies have used non-rectangular weighting strategies. A very recent example is Pedeli and Varin (2018) who have suggested non-rectangular weights of type

$$w_{ijg} = \begin{cases} K(\frac{t}{d+1}), 1 \leq t \leq d, \\ 0, \text{otherwise} \end{cases}$$

where $K(x)$ is described by Pedeli and Varin (2018) as a kernel functions with bounded support ($K(x) = 0$ for $x \geq 1$). Some examples of kernel weights are Triangular ($K(x) = 1 - x$), Epanechnikov or Quartic. Non-rectangular weights allow to decrease the contribution to the likelihood of gradually distant pairs with more flexibility than rectangular weights. As previously mentioned, few studies have departed from the selection routine outlined above. Bevilacqua et al. (2012) report that non-rectangular weights improve the efficiency of maximum pairwise likelihood estimators while Pedeli and Varin (2018) do not report any gains in effectiveness. More evidences are needed. In this paper, we propose to address this question and also introduce different weighting strategies. Our approach is based on the use of decay functions, which, in the context of this research work essentially the same way as kernel functions but are more general.

Decay functions have been studied and used in many various fields in order to model a data value that is decreasing over time: the population decline of colonies of animals, the half-life and decay of radioactive materials or the decreasing effectiveness of vaccines over time for example. There are many types of decay functions. Using decay functions to choose a weighting strategy for pairwise composite marginal likelihood estimation is not new. The full-pairwise likelihood approach as well as the pairwise likelihood of order d with rectangular weights are examples of decay functions. Indeed, the full-pairwise approach simply corresponds to a constant decay approach where there is no decay, ever, while the pairwise-likelihood of order d corresponds to a step-decay function where there is no decay for $i \leq d$ and then full decay ($w_{ijg} = 0$) at time d (where d corresponds to a description of the rate of decay, either the time until full-decay or the time until half-decay). As previously mentioned, the different kernels used by Pedeli and Varin (2018) are also a specific type of decay functions. Many different other functions accommodate non-rectangular weights. In this paper, we base many of the proposed weighting strategies on the Weibull function, which yields weights such as

$$w_{ijg} = e^{(-(\frac{t}{d})^k log(2))} \tag{8}$$

where k is a positive parameter which regulates the shape of the decay function and t corresponds to the (positive) value of the (temporal) distance between the two observations composing a pair (expressed in the same unit as d). It is worth noting that the Weibull decay function corresponds to the popular exponential decay function when $k = 1$. Figure 1 below provides a graphical representation of the Weibull function as well as the step-decay function and the constant decay approach.

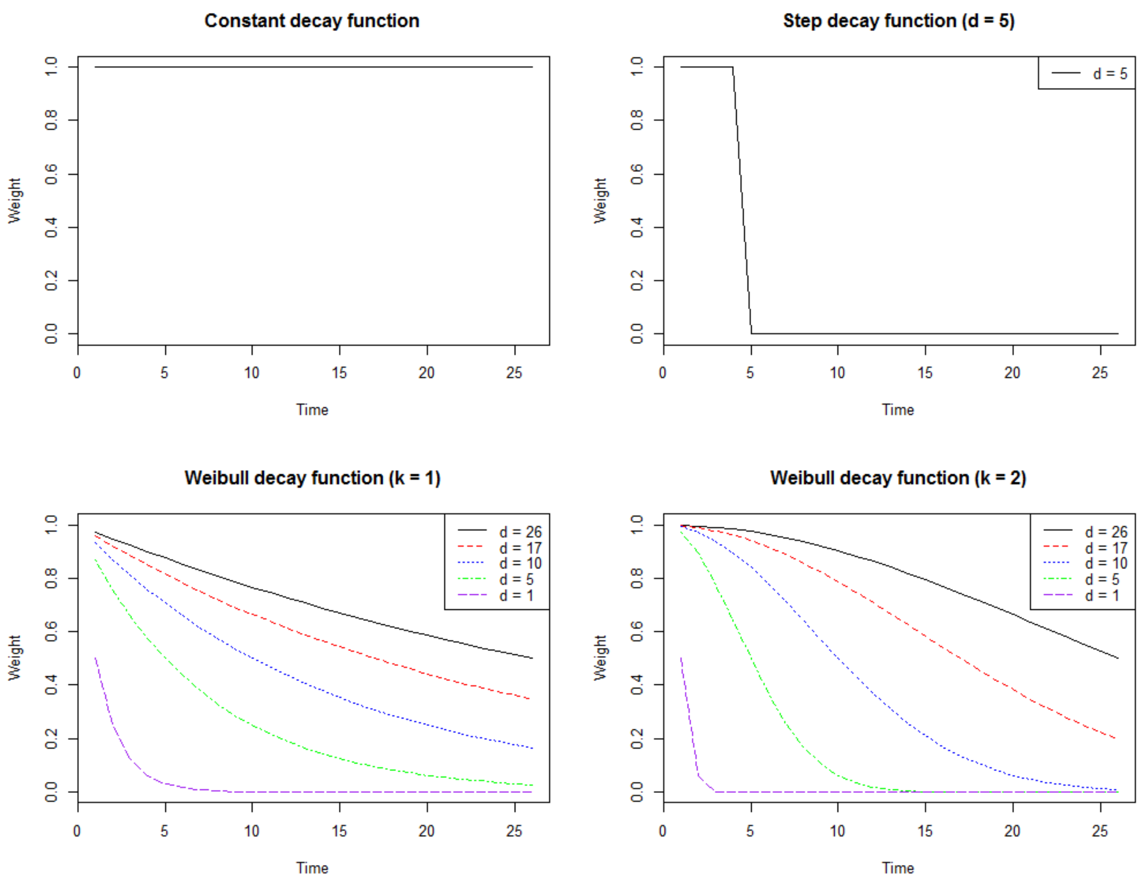

Figure 1: Main decay functions and corresponding weights

3.2.1 Heterogeneity

3.2.1.1 Random heterogeneity

Heterogeneity in decay functions can be introduced *via* various means. As previously observed, different values of d leads to different decay rates and hence to different weights. Moreover, the Weibull function can also yield different weighting structures by using different values for k. Finally, and more importantly, non-rectangular decay functions can be transformed into random non-rectangular decay functions by replacing t with $\frac{t}{s}$, where s is a random disturbance. As an illustration, Figure 2 below shows the weights derived from a Weibull decay function for which $d = 10$, k=2, $max(t) = 26$ and $s \sim LN(-1, 0.5)$. We label such an approach as random decay weighting, and we compare it with a completely random decay approach (labelled as random weighting) where weights are randomly distributed with $w_i \sim U(0, 1)$ for all i.

Random non-rectangular weights have never been used in the context of pairwise

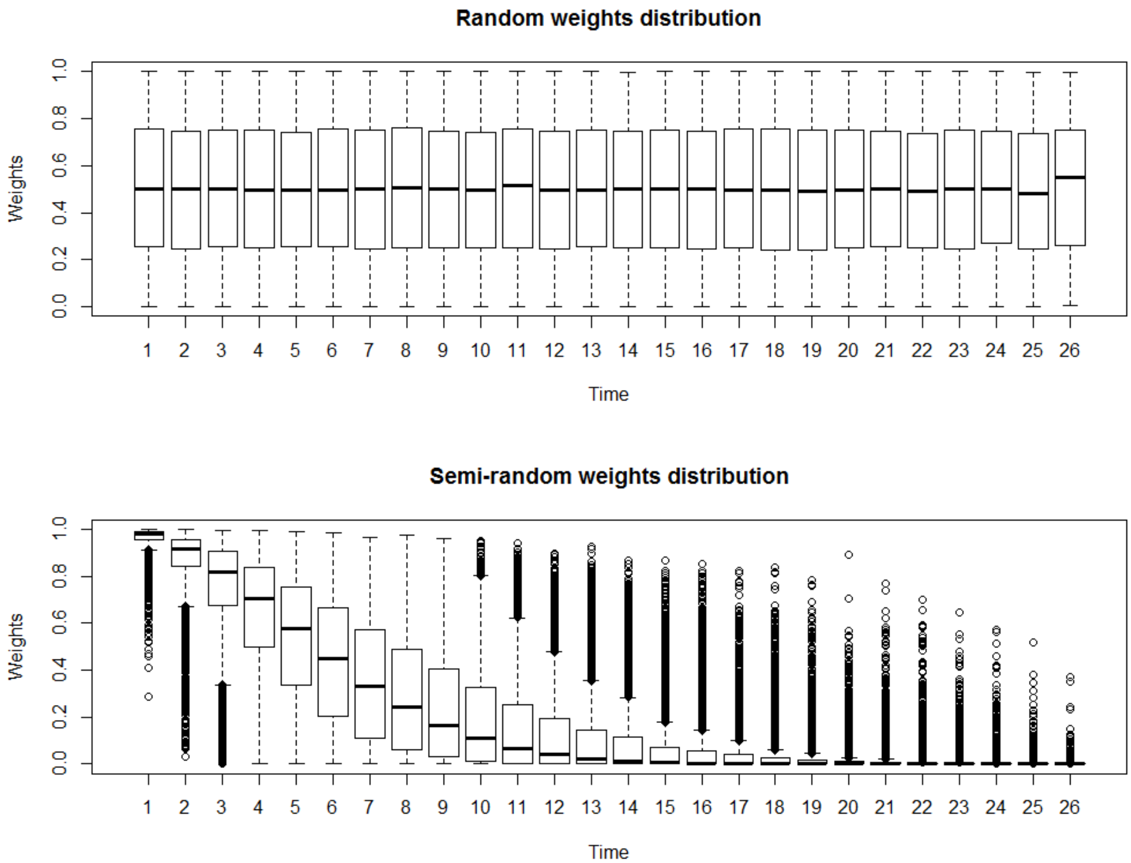

Figure 2: Random weighting *versus* random decay weighting - Illustration

composite marginal likelihood estimation to the best of our knowledge. Introducing a random disturbance in the decay function means that, from the point of view of the analyst, it is not sure whether a given pair of observations which are distant by 2 units, for example, are less informative than a pair of observations which are distant by 3 units or more. Figure 2 shows that on average, pairs are assumed to become less informative when distance increase, but it is hypothesised that it is not necessarily the case for all pairs and all individuals. The effect of allowing non-rectangular weights to vary randomly on model performances is assessed in the simulation study carried in the remainder of this paper.

3.2.1.2 Observed heterogeneity

Pairs of observations which are closer in time are likely to be more similar than pairs of observations which are distant in time. For example, a large range of studies has shown that individuals are less likely to have a car available for them to use when they just reach the legal age for driving than when are older (Beige and Axhausen, 2007; Oakil et al., 2014). In such a context, pairs of observations which are distant in time are more likely to report transitions in the life course (switching from one car availability status to another) than pairs of observations which are closer in time. Using the same sampling strategy regardless of whether the two elements of the pairs are stable or whether they indicate a transition in the life course might affect results because it might lead to underweight transitions, thus leading to biased estimates. In this paper, we explore whether using a different weighting strategy for pairs which feature a transition *i.e.* ($y_{ij} \neq y_{ig}$) and pairs which feature a stable behaviour ($y_{ij} = y_{ig}$) leads to better results. For example, one could specify

$$w_{ijg} = \begin{cases} e^{(-(\frac{t}{d})^2 log(2))}, y_{ij} \neq y_{ig}, \\ e^{(-(\frac{t}{d})^1 log(2))}, \text{ otherwise} \end{cases}$$

where it is also perfectly possible to specify different decay functions for pairs which features transitions and pairs which feature stable behaviour while also introducing random heterogeneity.

3.3 Random and semi-random pairs selection approaches

3.3.0.1 Random pairs

Another weighting strategy worth exploring is the case where weights are considered to be either 0 or 1 (*i.e.* pairs which enter the pairwise composite marginal likelihood functions all have the same weights, but not all pairs enter the function), but the selection is not only based on distance like it is the case with the step-decay function. We label the most simple case the random pairs selection approach, where the analyst specifies the number of pairs she wants to select from the total pool of pairs available and only selects a completely random subset for each individual in the dataset. For example, let's assume a dataset where each individual is observed once per year during 27 years. The total number of possible pairs is 351. The analyst may decide to only use C pairs out of 351 based on computational constraints or behavioural assumptions. One issue with such an approach is that it is likely to be very inefficient because in increases the chances of selecting pairs which are very distant apart and hence, potentially very uninformative.

3.3.0.2 Semi-random pairs

An alternative to a completely random pairs selection approach is to use one of the random decay functions previously introduced to select random but potentially informative pairs. More precisely, the analyst can generate weights for all the individuals in the considered dataset and all the pairs in a similar way to what has been previously described, and, for each individual, only select the C pairs for which the random weights are the highest and reassign a weight of 1 to these while the weights of all the remaining pairs are set to 0. Once again, it is possible to assign different decay functions to different pairs based on the problem at hand. The random and semi-random pairs selection approaches are illustrated in Figure 3, where again we assume a dataset where each individual is observed once per year during 27 years.

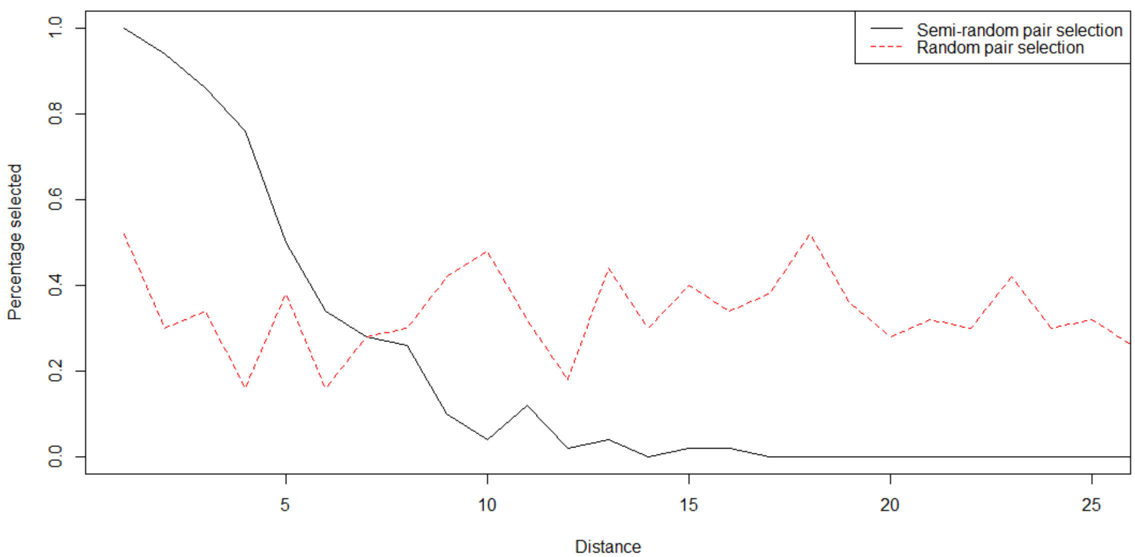

Figure 3: Random and semi-random pair selection, ($C = 120$)

3.4 A taxonomy of weighting strategies

To conclude and clarify this section, we provide a taxonomy of the different (existing and new) weighting strategies discussed in this paper and elsewhere in the literature. See Table 1 below. It is worth noting that most of the different weighting strategies discussed can accommodate both observed and random heterogeneity in weights as previously defined. In the remainder of this paper, we introduce the data on car availability over the life course which we use to demonstrate the feasibility of some of the approaches presented[2] and compare them with the (largely dominant) step-decay approach.

[2] We do not test the performance of all the decay functions presented but focus our analysis on the Weibull decay function without loss of generality. We introduce the other functions in order to illustrate the many different approaches one can choose to use beyond the constant decay and the step-decay approaches.

Table 1: Taxonomy - Weighting strategies

Approach	**Weights value**	**Additional information**
Constant	1	no decay for t < d, full decay (0) at time d
Step-decay	0 or 1	no decay for t < d, full decay (0) at time d
Triangular kernel	$1-(\frac{t}{d+1})$	For t < d, otherwise 0
Epanechnikov kernel	$\frac{3}{4}(1-(\frac{t}{d+1}))^2$	For t < d, otherwise 0
Quartic kernel	$\frac{15}{16}(1-(\frac{t}{d+1})^2)^2$	For t < d, otherwise 0
Triweight kernel	$\frac{32}{35}(1-(\frac{t}{d+1})^2)^3$	For t < d, otherwise 0
Tricube kernel	$\frac{70}{81}(1-(\frac{t}{d+1})^3)^3$	For t < d, otherwise 0
Exponential decay	$e^{(-(\frac{t}{d})^1 log(2))}$	Never reaches 0
Weibull decay	$e^{(-(\frac{t}{d})^2 log(2))}$	Never reaches 0
Hill decay	$\frac{1}{1+(\frac{t}{d})^2}$	Never reaches 0
Smooth-compact decay	$e^{k-\frac{k}{1-(\frac{t}{d})^2}}$	For t < d, otherwise 0
Random pairs	0 or 1	Randomly select C pairs out of the total number of pairs
Semi-random pairs	0 or 1	Randomly select C pairs out of the total number of pairs using a decay or kernel function

4 Data

Car availability over the life course is a well suited topic to study and compare the performances of different weighting approaches given the wide scope for spurious state dependency. For example, Dargay (2001) reports that car ownership is clearly associated with habit and resistance to change and that it is difficult to abandon even if the economic consequences of having a car available may not work out favourably for the owner. These factors may be difficult to measure, especially in a panel setting.

It is also worth underlining that the use of autoregressive models for investigating spurious state dependence in car availability or car ownership status is not a new topic. However, this is the first time to our knowledge that a *areoprobit* model has been used to model mobility biography data as described in the next paragraph. Previous attempts to use autoregressive models for analysing car availability or car ownership have either focused on microeconomic panel data, merging different data registries or pseudo-panel data obtained from consumption surveys (see Dargay (2001)). Life course calendar data on the other hand have often been analysed by static random-effects models, thus ignoring the potential effects of state dependency. In this section, we describe the data collection effort and provide details on the data sample we use for modelling car availability status over the life course.

4.1 Survey design

The data originates from a retrospective survey which has been carried out since 2007 at the Department of Transport Planning of the TU Dortmund as an annual homework project for students. Since 2012, it has been part of the collaborative project "Mobility Biographies: A Life-Course Approach to Travel Behaviour and Residential Choice" and additional data has been collected in Frankfurt and Zurich. The survey collects data about the students, their parents and grandparents. The students 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 recruit another person, preferably of the same generation. The questionnaire, which is the same for every generation, contains a series of retrospective questions about residential and employment biography, travel behaviour and holiday trips, as well as socio-economic characteristics.

As the sample has a unique structure it is not possible to appraise its representativeness (see Erickson (1979) for problems with representativeness in similar surveys). The majority of the respondents live in Dortmund (North Rhine-Westphalia), one of the most densely populated regions of Germany. Furthermore, within the grandparent generation, a bias to female participants who live longer on the one hand and are also often younger and more communicative can be recognized (Scheiner et al., 2014). Finally, retrospective data especially collected for as long a period as the life course always bears the risk of the so-called memory bias which means an unintended or intentional bias of the autobiographic memory (Manzoni et al., 2010). However, the whole study focusses on mobility behaviour in the life-course, so that the results are not expected to be significantly affected by the 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 Data sample

In this study we chose to focus on the parents' generation because it provides the longest series of observations in a contemporary setting. Our study sample features 437 women and 27 observations per woman. The reason why we only focus on women is because we want to exclude individuals which are correlated with one another (*i.e.* members of the same household). In addition, a preliminary analysis of the data has shown that women transition more from one car availability status to another than men during their life course, which is a desirable feature given the scope of the current paper (analysing very stable life courses might not require to use alternative weighting techniques such as those we discuss since there would be little or no variability over time). Each observation relates to one individual and one year, giving 11,799 observations in total. We include all the individuals from the parents' generation who were at least 18 in 1980. For each year, participants were asked to report their car availability status and had to choose between four (ordered) responses: "Never", "Sometimes", "Often" or "Always"[3]. In full detail, the concept of car availability can extend to include the holding of a driving licence by

[3]In this paper, we have decided to merge "Often" and "Sometimes" in the same category ("Sometimes") because the number given for these categories have been found to be much smaller. Preliminary models have been estimated in order to assess the impact of this decision on model estimates. We found that this does not substantially affect conclusions, especially given the purpose of our paper.

the individual and the ownership, renting or leasing of a car by household members, their employers or others (e.g. parents of students). Even more widely, car availability can extend to the availability of a car in which the individual can travel as a passenger. These responses are subjective simplifications of a complex concept, but nevertheless they represent states that persist through time, whether because of long-term characteristics like orientation towards cars and the holding of a licence, or medium-term characteristics like car ownership.

Figure 4 shows that car availability tends to change over time. The first category, "Never", significantly decreases between 1980 and 1990 while "Sometimes" and most importantly "Always" increase. There is little evolution overall between 1990 and 2006 so one may expect autocorrelation to be high. Our study sample also features annual measures of distance from work location, education level, number of children, residential location, marital status, driving licence holding and home ownership status. These variables will be used as independent variables in the *areoprobit* models. Table 2 below provides descriptive statistics and clarifies how these variables have been coded.

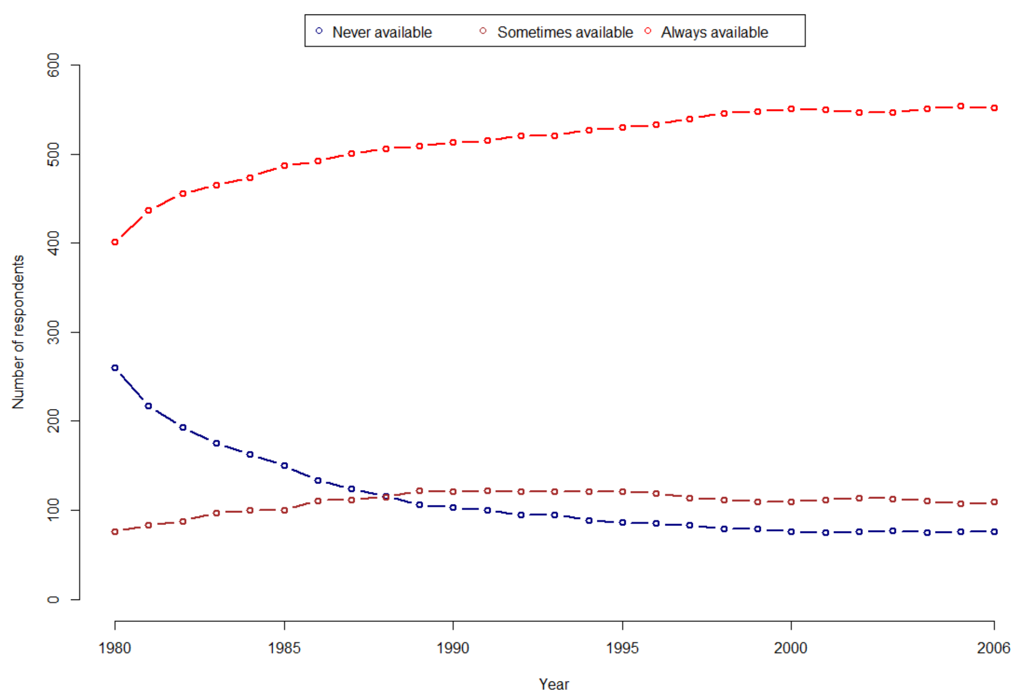

Figure 4: Evolution of the car availability status over the life course

Table 2: Descriptive statistics

Variable	Description	Mean	Std. dev.	Min.	Max.
car availability	0 if car is never available, 1 if sometimes available and 2 if always available	1.543	0.745	0	2
car_licence	1 if respondent has a driving licence	0.893	0.308	0	1
distw	Distance from work, in km	8.990	14.941	0	150
age	Age, in years	35.658	8.877	18	69
female	1 if respondent is female	0.631	0.482	0	1
own_home	1 if respondent owns her home	0.443	0.496	0	1
children	Number of children	1.511	1.135	0	9
married	1 if respondent is married	0.814	0.388	0	1
degree	1 if respondent has a university degree	0.294	0.455	0	1
big_city	1 if respondent lives in a city with population >300,000	0.449	0.497	0	1

5 Modelling work

The variables introduced in Table 2 are used as independent variables in a series of autoregressive random effects ordered probit models estimated *via* CML. We estimate two threshold parameters μ_1 and μ_2 (described in Equation (2)). Since the dependent variable, *car availability status*, can take one of three values for each individual and each year, it means that we normalise the constant to zero. We also estimate σ which corresponds to the random effect as previously defined. Finally, we estimate an autocorrelation parameter, ρ, which is bounded between 0 and 1 by means of a logistic transformation. We first estimate 25 models where pairs are selected using step-decay functions with d increasing for each model (from 2 to 26). As previously mentioned, the parameter estimates derived from the best step-decay model (based on the value of the trace of the robust variance-covariance matrix) will be used to set up our simulation study and compare the step-decay function approach to other strategies for selecting pairs.

5.1 Step decay models

Results for the step decay models are reported in Figure 7.

As previously discussed, we follow Bhat et al. (2010) and Varin and Czado (2009) and report the trace of the robust variance-covariance matrix for each model. The optimal d in this context corresponds to the model which minimises the trace. The simplest model only uses pairs which are not distant by more than two years (d=2), which corresponds to 51 pairs in total. The trace of the robust variance-covariance matrix is high (2.369), which suggests that it is necessary to move towards a model specification which makes use of more pairs. We find that increasing d decreases the trace. Unsurprisingly, we find mediocre values for the trace when the maximum distance between pairs is too low. The model which exhibits the best trace, 1.052 corresponds to $d = 17$. Finally, we find that the full-pairwise approach is not the most efficient model specification because it uses all the pairs (so it is computationally more burdensome), but reports a higher trace (1.112). This is a common result in the literature as indicated by Varin and Czado (2009), among others, and it confirms that a full-pairwise likelihood approach is not necessarily the best

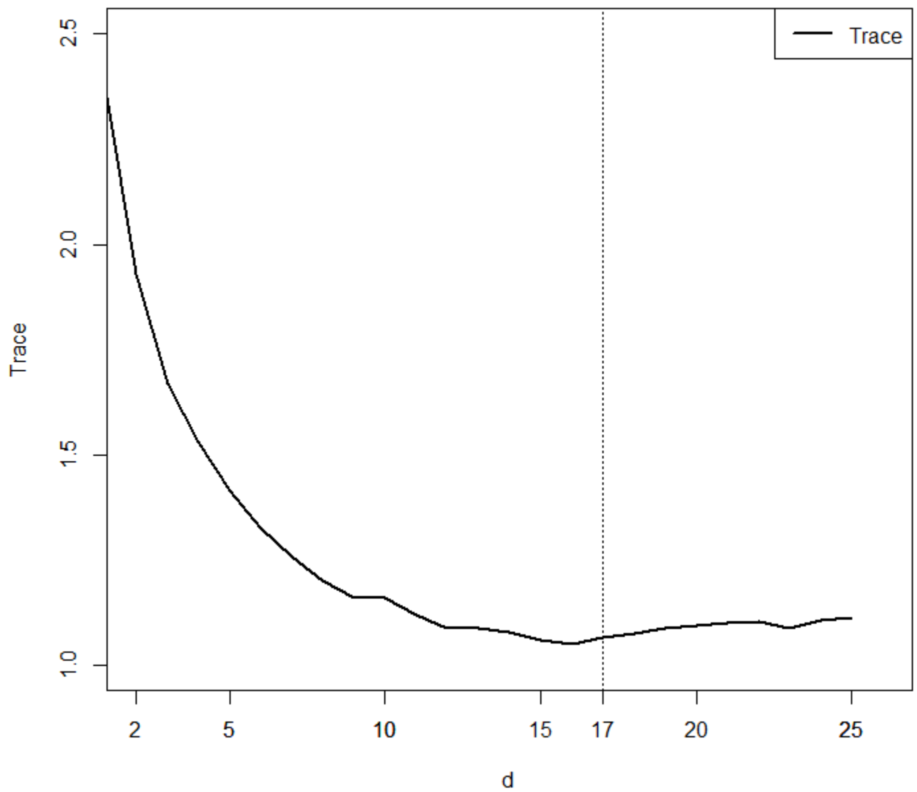

Figure 5: Step-decay model results

modelling strategy. Detailed outputs for the best step-decay models are provided below.

A rapid analysis of the model results indicates that all the parameters are significant and have the correct sign. In particular, having a driving licence is a strong predictor of car availability status. The value of the autocorrelation coefficient ρ is high (2.5413, which corresponds to 0.927 for the transformed parameter), which is not surprising: this result suggests that spurious state dependence has a strong influence on car availability status. In other words, ρ indicates that there are strong, constant unobserved factors which persistently influence our dependent variables across years. In the context of car availability, this means that previous car availability status, attitudes, lifestyles and habits have a strong influence on mobility behaviour. We now use these results to set up a simulation study in order to compare the finite-sample performance of the different weighting strategies previously introduced in this paper.

Table 3: Best step-decay model with $d = 17$

Trace	1.0527		
	Estimate	Rob.std.err.	Rob.T.
βage	0.1922	0.0362	5.31
βage^2	-0.2218	0.0465	-4.77
$\beta car\_licence$	2.5974	0.4766	5.45
$\beta distance\_work$	0.4173	0.0867	4.81
$\beta distance\_work^2$	-0.2966	0.0833	-3.56
$\mu 1$	4.5881	0.7686	5.97
$\mu 2$	0.2323	0.1630	1.43
σ	1.6414	0.3152	5.21
ρ	2.5413	0.3127	8.13

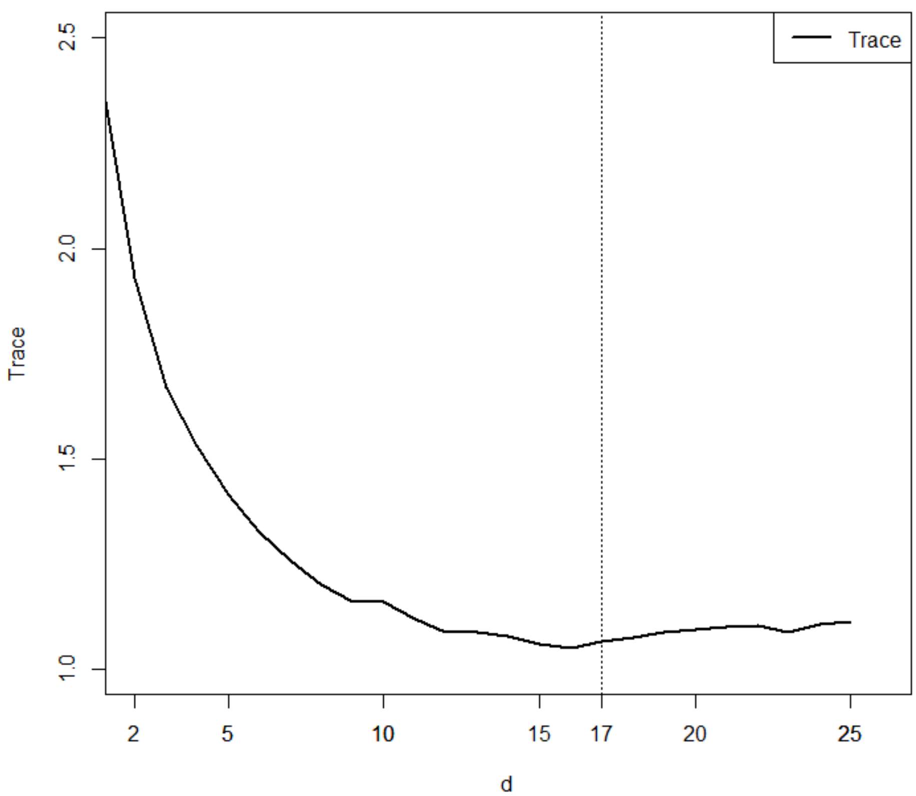

Figure 6: Trace results

5.2 Simulations

The objective of this simulation study is to compare the merits of the step decay approach approach (which is by far the most widely used in the literature) with other weighting strategies. We seek to generate a simulated dataset which is as close as possible to the real, actual dataset on car availability introduced in the previous sections of the present paper.

5.2.1 Data generation

The simulated dataset is generated as follows:

1. Let $y\_sim^*_{ij}$ the simulated continuous latent outcome from which we derive the simulated car availability level. See equation 1 for details.

2. We propose $y\_sim^*_{ij} = \beta\_sim' X_{ij} + \alpha\_sim_i + \epsilon\_sim_{ij}$, where we use the same exogenous variables as in the models previously estimated using real data. The vector of parameters β_sim as well as the variance of the random effects α and the autocorrelation ρ of the errors ϵ_sim_{ij} correspond to the values reported in Table 3. The value of the exogenous variables correspond, for each individual and each time period, to the values actually reported in the real dataset previously introduced in this paper. Autocorrelated errors are generated by following the recommendations of Paleti and Bhat (2013).

3. If the value of the simulated latent outcome is below the first threshold, the corresponding simulated car availability level is 0 (Never available) while if it is between the first and the second threshold, the corresponding simulated car availability level is 1 (Sometimes available) and 2 (Always available) else.

5.2.2 Sampling strategies

We simulate 100 datasets. For each dataset, we estimate the following models:

1. Step-decay models with d=5, 10 and 17.

2. Weibull decay models with k=2 and d=5, 10 and 17.

3. Weibull random decay models with k=2, d=17 and a disturbance $g \sim LN(-1, 0.5)$.

4. Random pairs models using 120, 215 and 306 pairs respectively (which corresponds to a computational burden which is equivalent to d=5, 10 and 17 in the context of a step-decay model). Weights are rectangular (0 or 1) and selected with $R \sim U(0, 1)$ and the method previously introduced in Section 3.

5. Semi-Random pairs models using 120, 215 and 306 pairs respectively. Weights are rectangular (0 or 1) and selected with R corresponding to a Weibull random decay with k=2, d=17 and a disturbance $g \sim LN(-1, 0.5)$.

6. Heterogeneous Weibull decay models with k=2 and d=17 for stable pairs and d=5 for transition pairs.

5.2.3 Performance measures

1400 models are estimated in total. We follow Paleti and Bhat (2013) and compute the following performance measures:

1. we compute the mean estimate for each model parameter across the 100 datasets as well as the bias for each model. This allows to calculate the Absolute Percentage Bias (APB) as well as the Root Mean Square Error (RMSE).

2. we then compute the standard deviation of the estimates of the parameter values across the 100 datasets (which is not the same as the asymptotic standard errors of the parameters, which are obtained by means of the sandwich estimator) and we label this the Finite Sample Standard Error (FSSE).

3. We calculate the mean of the asymptotic standard errors (obtained by means of the sandwich estimator) for each one of the 10 model parameters across the 100 models and the 14 sampling strategies and we name the resulting measures the Average Asymptotic Standard Errors (AASE).

4. Finally we compute the APB of the AASE for each parameter relative to the FSSE.

The results of this simulation study are provided in a series of tables and figures in the next subsection.

5.3 Simulation results

Sampling	Average Percentage Bias
Rand. pairs WB1	0.2645
WB 1	0.2829
Rand. pairs	0.2991
Step 26	0.1432
Triangular	0.1475
SC 1	0.3336
Hill	0.3396
Epanechikov	0.1530
W 0.5	0.3404
Quartic	0.1521
WB 2	0.3431
Tricube	0.1812
SC 0.5	0.3558
Step 17	0.1604
Triweight	0.1533
Step 10	0.2091
SC 2	0.4226
Step 5	0.3608

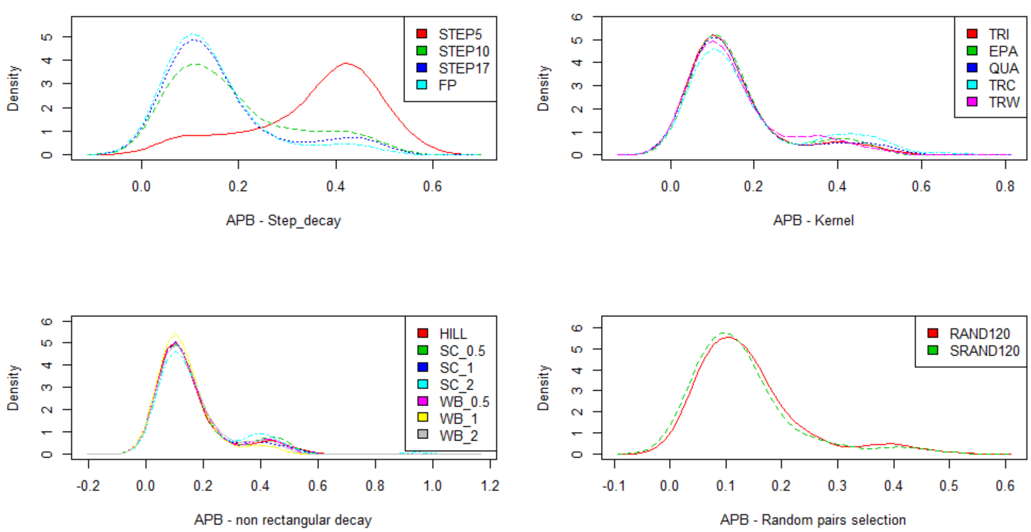

Figure 7: Simulation results

6 Further results

In this section we test the different approaches used in the simulation exercise using the real dataset. We compare the models using Bhat's predictive log-likelihood.

TO BE ADDED

7 Conclusions

This paper focuses on model specification and selection in the context of CML estimation. Using a unique data-set which covers 27 years of car availability data in Germany, we provide new insights on how to build a pairwise likelihood function for estimating AREOPROBIT models. In addition, we suggest the use of a new goodness-of-fit indicator, the ULL, based on the intractable MSL model which underlines CML specifications. Finally, we provide new results on spurious state dependence in car availability levels using one of the longest panel data-set found on this topic in the literature.

Our results suggest that CML functions built using a random selection of bivariate normal probabilities for each individual in the sample are more efficient than functions built using closest pairs. CML functions composed of random pairs provide better results in terms of ULL and are computationally much less burdensome. This important result reinforces the appeal of CML estimation methods in comparison with MSL estimation methods.

Our results also suggest that a purely random approach to pairs selection may not always be desirable. Indeed, the performances of a purely random pairs approach vary greatly depending on which pairs are randomly selected. As a result, we introduce smart pairs, which is a selection process where some pairs have more chances of being randomly selected than others, based on criteria set up by the researcher given the data at hand. The smart pair models introduced in this paper provide better results in terms of ULL and are more stable than purely random pair models. In other words, the smart pair models reduce the risks associated with using random pairs, that is to select uninformative pairs, while also improving model fit. The smart pairs approach proposed in this paper allows to obtain better models (in terms of goodness-of-fit and trace) at a fraction of the computational cost required for estimating a more complicated close pairs model or a full-pairwise model. indeed, we find that using 80 pairs out of 351 (more than 4 times less) allows to derive better results providing that the pairs are adequately chosen.

Finally, we explore the determinants of car availability over the life course with a specific focus on spurious state dependence. Our results strongly suggest that there exist unobserved factors which influence car availability and persist over years. We note that the life course calendar approach, which is the data collection method used in this paper, has many advantages over other survey methods but cannot recover essential information such as attitudes toward motorised mobility. As a result, we conclude that future efforts on this topic should consider new data collection approaches which can more adequately capture the information identified as spurious state dependence in our model.

The present study has some limits which should be acknowledged. Firstly, it only addresses the case of balanced panel datasets without missing values. Some adjustments may be necessary to apply the presented methodology in an unbalanced panel data context. More precisely, the selection of the smart pairs as well as the ULL measure may require modifications in such a context. Moreover, the bivariate normal probabilities may need to be weighted when using CML on an unbalanced panel (Joe and Lee, 2009; Kuk and Nott, 2000; Yi et al., 2011).

Future research on this topic should consider new ways of building CML functions. New criteria for selecting smart pairs and new weights should be tested. Random coefficient models built using smart pairs or random pairs are also worth investigating. Moreover, CML estimation allows us to use more complex autocorrelation structures which could provide more insights into how spurious state dependence evolves over time. For example, one could use a Toeplitz error structure instead of the classic AR(1) structure used in this paper and adapt the distance between pairs at different points in time based on how autocorrelation evolves over time. The CML approach is flexible and can be enriched in many ways as demonstrated by the recent stream of literature on this topic. Once again, how to build a CML function remains an open research area (Bhat et al., 2010).

8 Acknowledgements

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, 981–996.

Beige, S., Axhausen, K.W., 2007. The ownership of mobility tools during the life course .

Bevilacqua, M., Gaetan, C., Mateu, J., Porcu, E., 2012. Estimating space and space-time covariance functions for large data sets: a weighted composite likelihood approach. Journal of the American Statistical Association 107, 268–280.

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, 923–939.

Bhat, C.R., 2015. A new generalized heterogeneous data model (ghdm) to jointly model mixed types of dependent variables. Transportation Research Part B: Methodological 79, 50 – 77. doi:https://doi.org/10.1016/j.trb.2015.05.017.

Bhat, C.R., Castro, M., Khan, M., 2013. A new estimation approach for the multiple discrete–continuous probit (mdcp) choice model. Transportation Research Part B: Methodological 55, 1 – 22. doi:https://doi.org/10.1016/j.trb.2013.04.005.

Bhat, C.R., Pinjari, A.R., Dubey, S.K., Hamdi, A.S., 2016. On accommodating spatial interactions in a generalized heterogeneous data model (ghdm) of mixed types of dependent variables. Transportation Research Part B: Methodological 94, 240 – 263. doi:https://doi.org/10.1016/j.trb.2016.09.002.

Bhat, C.R., Sener, I.N., Eluru, N., 2010. A flexible spatially dependent discrete choice model: formulation and application to teenagers' weekday recreational activity participation. Transportation research part B: methodological 44, 903–921.

Bhat, C.R., Sidharthan, R., 2012. A new approach to specify and estimate non-normally mixed multinomial probit models. Transportation Research Part B: Methodological 46, 817 – 833.

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, 1–117.

Castro, M., Paleti, R., Bhat, C.R., 2012. A latent variable representation of count data models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections. Transportation Research Part B: Methodological 46, 253 – 272. doi:https://doi.org/10.1016/j.trb.2011.09.007.

Cherchi, E., Manca, F., 2011. Accounting for inertia in modal choices: some new evidence using a rp/sp dataset. Transportation 38, 679.

Clark, B., Lyons, G., Chatterjee, K., 2009. Understanding the dynamics of car ownership: Some unanswered questions .

Dargay, J., Hanly, M., 2007. Volatility of car ownership, commuting mode and time in the uk. Transportation Research Part A: Policy and Practice 41, 934–948.

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

Enam, A., Konduri, K.C., Pinjari, A.R., Eluru, N., 2018. An integrated choice and latent variable model for multiple discrete continuous choice kernels: Application exploring the association between day level moods and discretionary activity engagement choices. Journal of Choice Modelling 26, 80 – 100. doi:https://doi.org/10.1016/j.jocm.2017.07.003.

Erickson, B.H., 1979. Some problems of inference from chain data. Sociological methodology 10, 276–302.

Ferdous, N., Bhat, C.R., 2013. A spatial panel ordered-response model with application to the analysis of urban land-use development intensity patterns. Journal of Geographical Systems 15, 1–29. doi:10.1007/s10109-012-0165-0.

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

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

Hanly, M., Dargay, J., 2000. Car ownership in great britain: Panel data analysis. Transportation Research Record: Journal of the Transportation Research Board , 83–89.

Heckman, J.J., 1981. Heterogeneity and state dependence, in: Studies in labor markets. University of Chicago Press, pp. 91–140.

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

Hensher, D.A., Greene, W.H., 2003. The mixed logit model: the state of practice. Transportation 30, 133–176.

Joe, H., Lee, Y., 2009. On weighting of bivariate margins in pairwise likelihood. Journal of Multivariate Analysis 100, 670–685.

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

Kuk, A.Y., Nott, D.J., 2000. A pairwise likelihood approach to analyzing correlated binary data. Statistics & Probability Letters 47, 329–335.

Manski, C.F., Sherman, L., 1980. An empirical analysis of household choice among motor vehicles. Transportation Research Part A: General 14, 349 – 366. doi:https://doi.org/10.1016/0191-2607(80)90054-0.

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

Müggenburg, H., Busch-Geertsema, A., 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.

Narayanamoorthy, S., Paleti, R., Bhat, C.R., 2013. On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level. Transportation Research Part B: Methodological 55, 245 – 264. doi:https://doi.org/10.1016/j.trb.2013.07.004.

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

Paleti, R., Bhat, C., Pendyala, R., 2013. Integrated model of residential location, work location, vehicle ownership, and commute tour characteristics. Transportation Research Record: Journal of the Transportation Research Board , 162–172.

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

Papageorgiou, I., Moustaki, I., 2018. Sampling of pairs in pairwise likelihood estimation for latent variable models with categorical observed variables. Statistics and Computing doi:10.1007/s11222-018-9812-8.

Pauli, F., Racugno, W., Ventura, L., 2011. Bayesian composite marginal likelihoods. Statistica Sinica , 149–164.

Pedeli, X., Varin, C., 2018. Pairwise likelihood estimation of latent autoregressive count models. arXiv preprint arXiv:1805.10865 .

Sang, H., Genton, M.G., 2014. Tapered composite likelihood for spatial max-stable models. Spatial Statistics 8, 86 – 103. Spatial Statistics Miami.

Scheiner, J., Sicks, K., Holz-Rau, C., 2014. Generationsübergreifende mobilitätsbiografien–dokumentation der datengrundlage: Eine befragung unter studierenden, ihren eltern und großeltern. Dortmund, Deutschland: Technische Universität Dortmund .

Train, K., 2001. A comparison of hierarchical bayes and maximum simulated likelihood for mixed logit. University of California, Berkeley , 1–13.

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

Varin, C., Czado, C., 2009. A mixed autoregressive probit model for ordinal longitudinal data. Biostatistics 11, 127–138.

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

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

Yi, G.Y., Zeng, L., Cook, R.J., 2011. A robust pairwise likelihood method for incomplete longitudinal binary data arising in clusters. Canadian Journal of Statistics 39, 34–51.

Zhao, Y., Joe, H., 2005. Composite likelihood estimation in multivariate data analysis. Canadian Journal of Statistics 33, 335–356.