Heterogeneous hazard model of PEV users charging intervals: Analysis of four year charging transactions data

Seheon Kim a,⇑, Dujuan Yang b, Soora Rasouli a, Harry Timmermans a

a Urban Planning Group, Department of Urban Science and Systems, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
b Information Systems in the Built Environment, Department of Urban Science and Systems, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Abstract

Public charging infrastructure represents a key success factor in the promotion of plug-in electric vehicles (PEV). Given that a large initial investment is required for the widespread adoption of PEV, many studies have addressed the location choice problem for charging infrastructure using a priori simple assumptions. Ideally, however, identifying optimal locations of charging stations necessitates an understanding of charging behavior. Limited market penetration of PEV makes it difficult to grasp any regularities in charging behavior. Using a Dutch data set about four-years of charging transactions, this study presents a detailed analysis of inter-charging times. Recognizing that PEV users may exhibit different charging behavior, this study estimates a latent class hazard duration model, which accommodates duration dependence, unobserved heterogeneity and the effects of time-varying covariates. PEV users are endogenously classified into regular and random users by treating charging regularity as a latent variable. The paper provides valuable insights into the dynamics of charging behavior at public charging stations, and which strategies can be successfully used to improve the performance of public charging infrastructure.

1. Introduction

Several governments across the world have launched marketing campaigns and offered tax cuts to encourage people buying plug-in electric vehicles (PEV). These vehicles are viewed as an efficient and sustainable alternative to internal combustion engine vehicles (ICEV). Although the absolute market size of electric vehicles is still insignificant and consumer purchasing behavior seems to stagnate until further improvement in technology and competitive prices have become a reality (e.g., Rasouli and Timmermans, 2016), the market penetration of PEV raises issues about the availability of public charging infrastructure and its performance. Adequate availability of charging stations seems a sine qua non for the further acceptance and market penetration of electric cars.
The situation in the Netherlands can serve as an example. Tax reduction measures resulted in rapidly increasing sales of electric cars. Political parties have set the target of a substantially larger market share of electric cars by 2020. If these policy targets would be realized, a substantial investment in public charging infrastructure is required. Although the Netherlands is one of the leading countries in the provision of public charging stations (IEA, 2016), yet insufficient charging infrastructure and restricted driving range of PEVs are major barriers to large-scale adoption (Franke and Krems, 2013a). Consumers hesitate buying a PEV if not enough charging stations are available, whereas charging infrastructure providers are not willing to invest in new charging stations until there is a sufficient developed PEV market. This problem is a clear example of the so-called chicken-and-egg allegory (Kitamura and Sperling, 1987; Kang and Recker, 2014; Ko et al., 2016).

Further empirical studies are therefore warranted to enhance our understanding how charging behavior is embedded in daily activity-travel patterns. Such knowledge is needed to elaborate our models and improve the solutions offered to entice people purchasing more environmentally-friendly vehicles. Unfortunately, the present limited number of PEVs makes it hard to observe real-world charging behavior. Although various attempts have been made over the past few years to examine mobility and charging patterns of PEV users, most studies are based on small samples, often covering a short period of time (Franke and Krems, 2013b; Zoepf et al., 2013; Khoo et al., 2014; Speidel and Bräunl, 2014; Sun et al., 2015; Wen et al., 2016) or on Internal Combustion Engine Vehicle (ICEV) data based on an assumption that driving and refueling behavior would not be changed (Khan and Kockelman, 2012; Tamor et al., 2015; Wu et al., 2015; Jakobsson et al., 2016; Yang et al., 2016).

The aim of this study is to reduce this gap in our understanding of PEV users’ charging behavior at public charging infrastructure. Assuming that PEV users may differ in terms of charging frequency and regularity (regular charging versus random charging), we develop a hazard-based duration model of inter-charging times (i.e. the elapsed time between two consecutive charging events), to examine the regularity and frequency of charging events at public charging stations, using four-year longitudinal charging transactions data. Our empirical results enable segmenting PEV users into two groups with different charging regularity. The paper provides valuable insights into understanding the personal characteristics of frequent users, the frequency and amount of charging at public charging stations, and which strategies may be used successfully to improve the performance of public charging infrastructure.

The remainder of this paper is structured in the following manner. Section 2 reviews previous literature on charging behavior and inter-episode duration analysis. Section 3 describes the theoretical basis of the hazard-based duration model and the specifications used to model inter-charging times. Section 4 describes the data used in this study. Then, the results of model estimation, including the personal characteristics of regular and random users, are presented. The final section summarizes the main conclusions, implications of this study and plans for future research.

2. Literature review

2.1. Charging behavior

The study of charging behavior in transportation research started with the growing adoption of PEV around the world. Charging behavior studies mainly involved descriptive analyses of small samples or choice experiments based on hypothetical scenarios (Khan and Kockelman, 2012), Franke and Krems (2013b), for example, examined the charging patterns of 79 drivers in a 6-month EV trial in Berlin, focusing on concerns about the state-of-charge (SOC). Speidel and Bräunl (2014) examined the charging of 11 vehicles and usage patterns of 23 charging stations during the Western Australian Electric Vehicle Trial. In the Victorian EV Trial, Khoo et al. (2014) conducted a statistical analysis of the empirical relationships between vehicle/participant types and attributes of charging events including duration, charging time, daily charging frequency, energy consumed, and time to the next charging event. Zoepf et al. (2013) examined charging decisions of plug-in hybrid electric vehicle (PHEV) users at the end of each trip with a sample fleet of 125 cars across 1 year. Their results indicated significant heterogeneity in charging decisions. The Japan Automobile Research Institute (JARI) collected probe data from a larger sample of 483 BEVs over two years (Sun et al., 2015). Normal charging timing choice after the last trip of the day was analyzed using a mixed logit model. The probe installed in the vehicle provided rich information about charging behavior as well as vehicle information, such as vehicle trajectory and SOC. Wen et al. (2016) developed a latent class logit model for the charging choices of BEV drivers. Weldon et al. (2016) conducted a descriptive and statistical analysis to describe charging and trip making behavior of 72 BEV users in Ireland. Results suggested that BEV users tend to charge at regular intervals, which included both at-home charging and public charging, regardless of available range of their batteries. Despite the interesting results, it is hard to grasp the full picture of the usage of public charging infrastructure because the probes were only installed in BEVs. More recently, Daina et al. (2017) proposed a random utility model for jointly predicting EV drivers’ activity-travel scheduling and charging choices under smart charging scenarios. To develop the model, two series of stated choice experiments, each with 12 choice situations, were conducted among 88 respondents. The authors found significant heterogeneity across respondents regarding their preferences for charging schemes, considering desired SOC, charging duration and charging cost.

2.2. Inter-episode duration analysis

Several studies have identified the temporal structure underlying activity participation over a longer period of time (Axhausen et al., 2002; Bhat et al., 2005; Arentze and Timmermans, 2009; Rasouli and Timmermans, 2014). Particularly,
hazard models have been proposed to account for the dynamics of activity participation. Inter-episode duration analysis uses panel data, which has multiple observations per individual and focuses on the time elapsed since the last activity participation episode, called inter-episode duration or gap time, to examine the rhythmic structure of daily activities.

First, due to its recurrent nature, inter-episode duration analysis was applied in the study of shopping behavior by examining how marketing variables affect consumers’ decision to go shopping. Kahn and Schmittlein (1989) found evidence of weekly cycles in shopping trips by observing inter-shopping times using IRI (Information Resources, Inc.) shopping trip data, and classified shopping trips into “Quick” and “Regular” based on the amount of money spent. In the context of trip regularity, they provided insightful results by classifying consumers into two segments: those who made more quick trips than regular trips (i.e., Quicks) and those who made more regular trips than quick trips (i.e., Regulars). However, only one exogenous variable was used for the analysis of its impact on trip regularity. By taking duration dependence and heterogeneity in purchase rates across consumers into account, as well as time-varying covariate effects, Gupta (1991) developed four negative binomial distribution (NBD) models and compared these models. Results indicated that the model with duration dependence and heterogeneity in purchase rates improved overall model fit. Kim and Park (1997) also developed a NBD model of inter-shopping times using IRI shopping trip data, which incorporated heterogeneity in shopping regularity and frequency. Based on the parameter estimates, they endogenously segmented respondents into random and regular shoppers by taking the shape of the inter-shopping time distribution into consideration. They noted that the exponential distribution describes the inter-shopping time distribution of random shoppers, while the Erlang-2 distribution represents the distribution of shopping interval times of regular shoppers. Heterogeneity in shopping regularity was captured as a mixture of the exponential and the Erlang-2 distribution. Kim (2013) compared the pattern of online purchase timing with the corresponding patterns in the offline market. The results support the hypothesis that shopping time regularity collapses in the online market.

Secondly, a growing body of research has recently examined multi-day or multi-week longitudinal activity diary data (Yang and Timmermans, 2017; Langlois et al., 2016). Recognizing the limitation of reliance on cross-sectional data of existing activity-based models (Rasouli and Timmermans, 2014), previous research has stressed the necessity to examine temporal patterns in daily activity-travel behavior (Axhausen et al., 2002; Kitamura et al., 2006; Arentze and Timmermans, 2009). Previous research has applied inter-episode duration analysis to this problem. For example, Schönfelder and Axhausen (2001) used six-week travel diary data of 361 individuals from 162 households, located in the German cities of Halle/Saale and Karlsruhe, to analyze rhythms in “daily shopping” and “active sports” activities. They estimated the Cox proportional hazard model and the Weibull parametric hazard model for the inter-episode time distribution, which allows not only incorporating observed inter-episode durations for the probability of event occurrence, but also to treat duration-independent determinants such as socio-demographics or personal constraints. Bhat et al. (2004) used the same diary data to examine the regularity and frequency of shopping behavior. Based on a hazard-based duration model, they extended the model formulation to inter-episode duration analysis, accounting for a non-parametric baseline hazard and unobserved heterogeneity across individuals in inter-shopping durations, and applied a latent segmentation approach to segment shoppers into a random or a regular group according to their regularity in shopping activity participation. Bhat et al. (2005) extended the previous analysis to five different activity types, and found inherent weekly rhythms in activity participation. The results showed a stronger rhythmic pattern in non-shopping activities (i.e., social, recreation, and personal business activities) than in shopping activities (i.e., maintenance shopping and other shopping activities).

Our current study contributes methodologically and empirically to this literature on inter-episode duration analysis by elaborating the work of Kim and Park (1997) and Bhat et al. (2004). In terms of methodology, this paper extends a latent segmentation scheme for duration modeling to allow for the effect of time-varying covariates. Different modeling specifications for inter-charging times, which accommodate both observed and unobserved heterogeneity are tested to examine regularity in charging patterns across PEV users. Empirically, this paper deals with periodicity in charging transactions data, and segments PEV users into two broad groups according to charging regularity (i.e., random and regular). To the best of our knowledge, existing work on inter-charging times is largely restricted to descriptive statistics, and no predictive model has been developed using real-world charging data. The model results provide information to private and public sector agents for the future development of charging infrastructure and facilitate understanding the dynamic structure of time-to-charging events.

3. Model structure and specification

Before discussing the methodology underlying this study, it is useful to provide a definition of charging regularity. We posit that PEV users may be heterogeneous in terms of charging intervals. The differences in charging intervals between two hypothetical PEV users are presented in Fig. 1. Inter-charging times of regular users show little variance compared to those of random users. In other words, regular users charge their PEV at a relatively fixed interval, and thus it can be said that the timing depends on the time elapsed since last charging event. Thus, the concept of regularity relates to the existence of regularity in interval times between charging events. Note that individuals may exhibit a certain degree of variation themselves, implying that segmentations are probabilistic.

Since its introduction by Herniter (1971), the Erlang-\(k\) family of distributions has been extensively used to characterize the frequency distribution of the elapsed time between consecutive episodes in inter-episode duration analysis (Jeuland...
et al., 1980; Gupta, 1988, 1991; Kim and Park, 1997; Schweidel and Fader, 2009; Platzer and Reutterer, 2016). As a special case of the gamma distribution, where shape parameter $k$ is a positive integer, the Erlang-$k$ distribution allows for varying degrees of regularity in inter-episode duration. The higher $k$, the stronger an individual’s temporal regularity in activity participation. In other words, a high value for $k$ implies that the timing of the next episode is highly dependent on time elapsed since the last episode, whereas the Erlang-$k$ distribution can also be reduced to the exponential distribution if $k$ equals 1, which implies that the timing of the next episode is independent of the time elapsed since the last episode. As an illustrative example, Fig. 2 provides the frequency distributions of inter-charging times of the two users in Fig. 1. In this study, the exponential distribution ($k = 1$) is used to describe the memoryless property of random users as shown in Fig. 2(a). On the other hand, Fig. 2(b) shows a clear peak, which suggests the Erlang-$k$ distribution ($k > 1$) is a candidate distribution for regular users. However, the question remains: how regular is regular? An Erlang-2 timing distribution ($k = 2$) has a rich history in marketing research and has often served as a reference to account for the regular patterns in frequently occurring behavior, such as particular brand purchases (Chatfield and Goodhardt, 1973), product-category purchases (Jeuland et al., 1980; Lawrence, 1980; Dunn et al., 1983; Gupta, 1988, 1991; Fader et al., 2004; Schweidel and Fader, 2009), and shopping trips (Kim and Park, 1997).

Based on the findings of previous inter-episode duration analyses, some hypotheses can be formulated for this study. The first hypothesis to be tested is that individual heterogeneity in charging frequency and charging regularity exists among PEV users. To capture unobserved heterogeneity in charging frequency, a gamma mixing distribution is used, while a mixture of the exponential and Erlang-2 distribution is used to examine heterogeneity in charging regularity. We endogenously divide PEV owners into random and regular users using a latent segmentation approach by assuming that charging regularity is a latent variable, which is not directly measurable. The posterior membership probability of each user is obtained from the estimated parameters. It takes into account different degrees of regularity and randomness in an individual’s transactions data.

The second research hypothesis to be tested is that harsh weather conditions, which vary over time, decrease PEV users’ inter-charging time hazards (probability of charging occurrence since the last charging event). We realize that seasonal effects may induce people to use particular transport modes. However, inter-charging times concern shorter time periods. Harsh weather conditions may induce people to stay at home or delay charging to the next day if possible. This section describes the hazard-based duration model used to test the proposed hypotheses by examining inter-charging times.

### 3.1. Model for random users

Similar to previous studies, it is assumed that the exponential distribution is suitable to model inter-charging times of random users, while the Erlang-2 distribution is appropriate for modeling regular users. These properties can be easily illustrated by inspecting the distribution of inter-charging times. Let $t$ denote the inter-charging time between two successive charging events. If we suppose a random charging process for user $i$, without consideration of the covariates which will be discussed later, the probability density function $f^{\text{Exp}}_i(t)$ and survival function $s^{\text{Exp}}_i(t)$ of an exponential inter-charging time are equal to

$$f^{\text{Exp}}_i(t) = \lambda^{\text{Exp}}_i \exp(-\lambda^{\text{Exp}}_i t)$$

$$s^{\text{Exp}}_i(t) = \exp(-\lambda^{\text{Exp}}_i t)$$

where $\lambda^{\text{Exp}}_i$ is the inter-charging time hazard of user $i$.

The hazard function for exponential inter-charging times thus becomes $\lambda^{\text{Exp}}_i(t) = \frac{f^{\text{Exp}}_i(t)}{s^{\text{Exp}}_i(t)} = \lambda^{\text{Exp}}_i$, which is constant over inter-charging intervals. In other words, if an individual’s inter-charging time is distributed exponentially, his or her hazard rate of charging is independent of the time elapsed since the last charging event, which represents a complete random process of charging.
charging activity. Given that the charging transactions dataset is right-censored, the conditional likelihood function for random user \( i \) becomes

\[
L_i^{\text{Exp}}(t|\lambda_i^{\text{Exp}}) = \left[ \prod_{j=1}^{n(i)-1} f_i^{\text{Exp}}(t_j) \right] S_i^{\text{Exp}}(t_{m(i)}) = \left[ \prod_{j=1}^{n(i)-1} \lambda_i^{\text{Exp}} \exp(-\lambda_i^{\text{Exp}} t_j) \right] \exp(-\lambda_i^{\text{Exp}} t_{m(i)})
\]

where \( t_j \) is the \( j \)-th inter-charging time of user \( i \), \( n(i) \) is the last observation of user \( i \) (i.e., censored inter-charging time).

However, it is illogical to assume that all individuals in the concerned population have the same inter-charging time hazard. Unobserved heterogeneity in the inter-charging hazard can be introduced by assuming \( \lambda_i^{\text{Exp}} \)'s is gamma distributed across individuals. The gamma distribution is flexible since it does not impose any symmetry restrictions and takes on a variety of shapes depending on its parameters. Furthermore, the gamma distribution has often been employed for its algebraic simplicity and model parsimony. Therefore, unobserved individual heterogeneity can be introduced by allowing \( \lambda_i^{\text{Exp}} \) to be gamma distributed as follows.

\[
L_i^{\text{Exp}}(t|\lambda_i^{\text{Exp}}, \gamma_i^{\text{Exp}}) = \int_0^\infty L_i^{\text{Exp}}(t|\lambda_i^{\text{Exp}}) g(\lambda_i^{\text{Exp}}|\lambda_i^{\text{Exp}}, \gamma_i^{\text{Exp}}) d\lambda_i^{\text{Exp}}
\]

where \( g(\lambda_i^{\text{Exp}}|\lambda_i^{\text{Exp}}, \gamma_i^{\text{Exp}}) \) is the density function of the gamma distribution with scale parameter \( \lambda_i^{\text{Exp}} \) and shape parameter \( \gamma_i^{\text{Exp}} \).

This specification has advantages over the usual way of including heterogeneity in proportional hazard models (Lancaster, 1990). It provides a closed-form likelihood expression and allows one to interpret the degree of heterogeneity in the data in terms of the gamma distribution parameters.

3.2. Model for regular users

The Erlang-2 model is duration dependent, suitable for describing the charging pattern of regular users. This model is based on the assumption that the elapsed time since the last charging event may affect the decision whether or not to charge. Without consideration of the covariates, the probability density function \( f_i^{\text{Er}}(t) \) and survival function \( s_i^{\text{Er}}(t) \) of Erlang-2 can be written as follows:

\[
f_i^{\text{Er}}(t) = \left( \lambda_i^{\text{Er}} \right)^2 t \exp(-\lambda_i^{\text{Er}} t)
\]

\[
s_i^{\text{Er}}(t) = (1 + \lambda_i^{\text{Er}} t) \exp(-\lambda_i^{\text{Er}} t)
\]

Thus, the hazard function of the Erlang-2 model is \( \lambda_i^{\text{Er}}(t) = \frac{f_i^{\text{Er}}(t)}{s_i^{\text{Er}}(t)} = \frac{\lambda_i^{\text{Er}} t}{\left(1 + \lambda_i^{\text{Er}} t \right)} \). This equation shows that the hazard is monotonically increasing in duration and reflects positive duration dependence. Therefore, this model can be used to represent regular charging behavior. The conditional likelihood function of user \( i \) can be written as

\[
L_i^{\text{Er}}(t|\lambda_i^{\text{Er}}) = \left[ \prod_{j=1}^{n(i)-1} f_i^{\text{Er}}(t_j) \right] s_i^{\text{Er}}(t_{m(i)}) = \left[ \prod_{j=1}^{n(i)-1} \left( \lambda_i^{\text{Er}} t_j \right) \exp(-\lambda_i^{\text{Er}} t_j) \right] \left(1 + \lambda_i^{\text{Er}} t_{m(i)} \right) \exp(-\lambda_i^{\text{Er}} t_{m(i)})
\]

where \( t_j \) is the \( j \)-th inter-charging time of user \( i \), \( n(i) \) is the last observation of user \( i \).

Similar to Eq. (4), it is assumed that the \( \lambda_i^{\text{Er}} \)'s are gamma distributed to accommodate unobserved heterogeneity across individuals. The likelihood function of user \( i \) can be rewritten as

\[
L_i^{\text{Er}}(t|\lambda_i^{\text{Er}}, \gamma_i^{\text{Er}}) = \int_0^\infty L_i^{\text{Er}}(t|\lambda_i^{\text{Er}}) g(\lambda_i^{\text{Er}}|\lambda_i^{\text{Er}}, \gamma_i^{\text{Er}}) d\lambda_i^{\text{Er}}
\]
where \( g(\lambda^E_2|\beta_{E2}, \gamma_{E2}) \) is a density function of the gamma distribution with scale parameter \( \beta_{E2} \) and shape parameter \( \gamma_{E2} \).

### 3.3. Time-varying covariates

Thus far, we have expressed the likelihood function without consideration of the effects of covariates. It is, however, of interest to examine how the inter-charging hazard varies in response to a set of covariates. A proportional hazard approach is a well-known method, which allows the covariate effects to be multiplicative on the baseline hazard as follows (Wooldridge, 2010).\(^3\)

\[
\lambda_i(t) = \lambda_0 \exp(\mathbf{x}_i \beta)
\]  

(9)

where \( \lambda_0 \) is a baseline hazard which is assumed to be time-invariant, \( \mathbf{x}_i \) is a row vector of covariates, and \( \kappa(\cdot) \) is a non-negative function of \( \mathbf{x}_i \). The baseline hazard, \( \lambda_0 \), is the same across all individuals in the concerned population, and the individual hazard function, \( \lambda_i(t) \), is shifted proportionally to a function \( \kappa(\mathbf{x}_i) \) of covariates. Due to the non-negativity assumption and ease of specification, an exponential function is usually used as follows, \( \kappa(\mathbf{x}_i) = \exp(\mathbf{x}_i \beta) \), where \( \beta \) is a column vector of coefficients. In this case, a 1% change in a covariate \( \mathbf{x}_i \) shifts the hazard function by \( \mathbf{x}_i \beta \).%

It is to be noted that Eq. (9) only holds for time-invariant covariates. A proportional hazard specification\(^4\) with time-varying covariates needs more careful consideration. In a proportional hazard specification, time-varying covariates have to meet the exogeneity condition (Lancaster, 1990). Any time-varying covariate whose path is determined independently of whether any person has or has not left the state in question (i.e., occurrence of charging event in a model for inter-charging times) is exogenous. Kalbfleisch and Prentice (2002) referred to this class of covariates as external covariates. An example of such a covariate might be weather whose path is presumably independent of any particular person charges his or her PEV or not. In contrast, the exogeneity condition would be violated if the path of a covariate is not defined once the person left the state. Lancaster (1990) gives the example of the wage paid in a job tenure duration, because it is not possible to define future wage if someone leaves the job. Kalbfleisch and Prentice (2002) named these as internal covariates. Lastly, there is the case that the covariate is not defined externally but it is not necessarily endogenous (i.e., path of covariate is still defined after the person leaves the state). An example might be state-of-charge (SOC) of PEV in a model for inter-charging times. Undoubtedly, SOC depends on the occurrence of a charging event, but its path is well-defined after charging. Thus, SOC could be either endogenous or exogenous for inter-charging times. It is an empirical issue whether such a covariate satisfies the exogeneity condition. In the proportional hazard specification, it is difficult to relax the exogeneity assumption with consideration of both time-varying covariates and unobserved heterogeneity together (Wooldridge, 2010). One possible way to deal with this problem is to discretize time, such as a day or a week, and to specify hazard as a piece-wise constant within time interval \( k \), \([ak-1, ak]\), where the time-varying covariates are also constant. In other words, hazard rates and covariates are constant within each time interval but may change over time intervals. If the current charging is made in time interval \( k \), by individual \( i \), the hazard rates can be written as

\[
\lambda_{ik}(t) = \lambda_0 \exp(\mathbf{x}_{ik} \beta), \quad a_{k-1} < t < ak
\]  

(10)

where \( k \) is the time interval when the current charging is made, \( k = 1, \ldots, K \), given that 1 is that of previous charging. With time-varying covariates, Eqs. (1), (2), (5), and (6) can be rewritten as follows (see Gupta, 1991).

\[
\int_{0}^{\lambda_{ik}(t)} \lambda_0 \exp(\mathbf{x}_{ik} \beta) - \lambda_0 \exp(\mathbf{x}_{ik} \beta) dt = \lambda_0 \exp(\mathbf{x}_{ik} \beta) - \lambda_0 \exp(\mathbf{x}_{ik} \beta) dt
\]  

(11)

and

\[
\int_{0}^{\lambda_{ik}(t)} \lambda_0 \exp(\mathbf{x}_{ik} \beta) - \lambda_0 \exp(\mathbf{x}_{ik} \beta) dt = [1 + \mathbf{0}_k(t)] \exp(\mathbf{x}_{ik} \beta) - \lambda_0 \exp(\mathbf{x}_{ik} \beta) dt
\]  

(14)

where \( \mathbf{0}_k(t) = \int_{0}^{t} \lambda_0 (s) ds \) is a column vector of coefficient for exponential model, and \( \beta_{E2} \) is a column vector of coefficient for Erlang-2 model. Like the preceding case in which we introduced unobserved heterogeneity into the models by allowing \( \lambda^E_1 \) and \( \lambda^E_2 \) to be gamma distributed, we allow \( \lambda_0 \) to have a gamma distribution for the model with time-varying covariates.

### 3.4. Latent segmentation scheme

However, thus far there is no information on charging regularity as it is considered a latent variable. In this paper, the propensity to belong to a segment (e.g., random user vs. regular user) is calculated using the posterior segment membership
probability of user $i$. The final unconditional log-likelihood function for $N$ users can be obtained by integrating the two likelihoods, conditional on the user being a random user or a regular user, weighted by $\varphi$ and $1 - \varphi$ (Kim and Park, 1997).

$$LL = \sum_{i=1}^{N} \log L_i(t|\beta, \varphi, \alpha_{\text{Exp}}, \gamma_{\text{Exp}}, \alpha_{\text{Exp}2}, \gamma_{\text{Exp}2}) = \sum_{i=1}^{N} \log \left[ \varphi L_{i\text{Exp}}(t|\beta, \alpha_{\text{Exp}}, \gamma_{\text{Exp}}) + (1 - \varphi) L_{i\text{Exp}2}(t|\beta, \alpha_{\text{Exp}2}, \gamma_{\text{Exp}2}) \right]$$

(15)

This log-likelihood can be maximized to estimate the following parameters: (a) scale parameter $\alpha_{\text{Exp}}$ and shape parameter $\gamma_{\text{Exp}}$, for the random user segment, (b) scale parameter $\alpha_{\text{Exp}2}$ and shape parameter $\gamma_{\text{Exp}2}$, for the regular user segment, and (c) weight $\varphi$ and the vector $\beta$.

The log-likelihood functions were maximized using optimx in R statistical software (Nash and Varadhan, 2011). Once parameter values are obtained, each individual can be assigned to the segments. The probabilities of assigning a given PEV user to one of the segments can be calculated through an empirical Bayes procedure. The posterior probability of user $i$’s membership in the random user segment can be calculated as $\varphi L_{i\text{Exp}} / L$, while the corresponding probability to be a regular user equals $(1 - \varphi) L_{i\text{Exp}2} / L$.

4. Data

4.1. Charging transaction data

This study is based on charging transactions data provided by E-laad, which is one of the largest foundations in the Netherlands managing public and semi-public charging infrastructure. The data contains information about normal charging, also referred to as slow charging, at 1880 charging stations between 2010 and 2014. Fig. 3 shows the spatial coverage of the data, indicating that the location and the usage of charging stations are concentrated in heavily populated areas. The charging transactions dataset consists of two parts: charging station information and transaction information. Charging station information includes data regarding the charging infrastructure such as charge station ID, address, postal code, and coordinates (longitude, latitude) of charging stations, while transaction information contains data regarding charging transaction ID, charging pass card ID, charging start time, charging end time, and plug off time.

Each charging transaction is linked with a charging pass card. It has to be noted we assume that one charging pass card is owned and used by a single user, even though it may be shared with others. Therefore, we assume that repetitive use of a charging pass card ID represents the charging pattern of the same individual. Unfortunately, the data used in this study does not provide socio-demographic data. However, it includes information about type of EV (i.e., battery electric vehicle vs. plug-in hybrid electric vehicle) for every transaction, which was detected based on different energy consumption profiles of vehicles types.

The proposed approach is applied to the charging transactions panel data. The analysis of this data has the following challenges. First, the data contains not only ordinary charging transactions, but also abnormal charging transactions. It is observed that some charging transactions occur too many times a day, and are largely associated with specific individuals. We considered these data as not relevant for our purpose, and deleted 12 individuals who charged more than 20 times a day. Second, in order to obtain stable model results, very light users (i.e., users who do not charge at least five times) were also eliminated. Third, some inter-charging times were too large (e.g., 1263 days), which may occur for many different reasons. Despite the E-laad data covers a significant number of charging stations in the Netherlands, there are charging stations that do not belong to the E-laad network. We assumed that PEV users in the data do not use a charging station outside the sample. It is also possible that users may be out of town for a long time. These extremely long inter-charging times, which are higher than 30 days, are considered as outliers, and were eliminated from the analysis. The upper bound was used here due to the small number of inter-charging times exceeding this duration length (about 7% of the observations). The sample used in this study consists of the inter-charging times of 9027 PEV users. Table 1 shows the summary statistics of the data used. The number of inter-charging duration spells ranged between 5 and 930, with an average of 48.83. The average inter-charging time of all individuals is 5.36 days.

The first step to understand the characteristics of inter-charging times is to examine the frequency distribution. Fig. 4(a) represents the frequency distribution of inter-charging times at an aggregate level, and resembles an exponential distribution. It may suggest that the EV users’ propensity to charge at time $t$ is independent of the elapsed time since the last charging event. Thus it seems that charging occurs randomly rather than regularly at first glance. However, Fig. 4(b) provides clear visual evidence of rhythmic patterns in charging activity. The sample hazard was calculated using the Kaplan–Meier method with an upper bound of 30 days. The Kaplan–Meier estimator or sample hazard is a non-parametric estimator of the hazard function that is calculated as the number of terminated episodes in discrete time period $k$ divided by the population still in the risk set in that period. Therefore, at an aggregate level, the Kaplan–Meier method can be used to check the underlying distribution of the data without controlling for observed and unobserved heterogeneity across individuals. The results are consistent with our contention that several peaks in each week are related to the charging patterns of regular users. On the other hand, the high hazard at the first 6 days is monotonically decreasing indicating negative duration dependence, which represents the inter-charging times of random users. In Section 5, we will discuss how random users and regular users are separated out and compared.
Fig. 3. Distribution of the charging stations.

Table 1
Summary of the charging transaction dataset.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Time span</td>
<td>8/20/2010–8/29/2014</td>
</tr>
<tr>
<td>Total number of users</td>
<td>9027</td>
</tr>
<tr>
<td>Total number of inter-charging times</td>
<td>449,844</td>
</tr>
<tr>
<td>Number of inter-charging times per person</td>
<td>Min = 5, Max = 930, Mean = 48.83</td>
</tr>
<tr>
<td>Length of inter-charging time (day)</td>
<td>Min = 0.01, Max = 29.99, Mean = 5.36</td>
</tr>
</tbody>
</table>

Fig. 4. (a) Frequency distribution of inter-charging times and (b) sample hazard.
4.2. Weather data

In this study, we focus on the influence of weather conditions on PEV users’ charging decisions. The weather data used were derived from the Royal Netherlands Meteorological Institute (KNMI) database. The database consists of daily values of weather variables such as precipitation, temperature, wind speed, etc., measured at 35 weather stations across the Netherlands. Charging transaction data were spatially and temporally linked to daily weather data. Each charging transaction record was augmented with weather information from the closest weather station on the corresponding day. We expect that harsh weather conditions may trigger PEV users to postpone charging at public charging stations to the next day or later, which can be incorporated into the model by means of the effects of covariates on the hazard rate (Böcker et al., 2016).

The variables used in the analysis are listed in Table 2. Four weather variables are converted into dummy variables taking the value of 1 if the corresponding variable reflects unpleasant weather conditions for charging outside. Thus, all variables considered here are expected to decrease users’ charging rates.

5. Empirical results

5.1. Model estimation results

Table 3 shows the estimation results for the different model specifications. First, the models with and without heterogeneity in charging frequency are compared. It is clearly shown that the model with gamma heterogeneity improves model fit in all cases. Given that the magnitude of the shape parameter indicates the degree of heterogeneity in the sample, the low value of the shape parameter indicates that substantial heterogeneity exists among individuals with respect to inter-charging times, which is in line with findings of other charging behavior studies (e.g., Zoepf et al., 2013).

Second, we assess the impact of including duration dependence in the model by comparing Exponential/gamma with Erlang-2/gamma models. The model without duration dependence fits the data much better. In other words, the model with the memoryless property of the exponential distribution describes the charging behavior of the entire set of observations well, which implies that PEV users tend to charge their cars more at random. In contrast to the results of previous inter-purchase and inter-shopping time studies (Gupta, 1991; Bhat et al., 2004), these results suggest that charging behavior has a different underlying structure than shopping behavior in terms of regularity. In addition, the interpretation of the shape parameter provides useful insight in the understanding of unobserved heterogeneity. While the coefficient of variation (CV) is a direct measure of the heterogeneity, the shape parameter (γ) is inversely proportional to it, as CV = \frac{\text{Standard deviation}}{\text{Mean}} = \frac{\sqrt{\frac{\pi^2}{6}}}{\frac{\gamma}{2}} = 1/\sqrt{\gamma}. Therefore, the lower the value of γ, the higher the heterogeneity. When comparing Exponential/gamma with Erlang-2/gamma models, the results show that the shape parameter for the latter model is marginally smaller than the shape parameter of the former model. This result has an intuitive interpretation based on the assumption that the total variance in PEV users’ inter-charging times consists of within-individual variance (i.e., intrapersonal variability) and between-individual variance (i.e., interpersonal variability). As we noted earlier, the Erlang-2/gamma model assumes an Erlang-2 distribution for inter-charging times of individuals. This assumption implies that charging occurs more regularly than in the exponential model, which can be interpreted as the Erlang-2 model (or Erlang-2/gamma) having a smaller within-individual variance than the exponential model (or Exponential/gamma). Thus, given the same amount of total variance in inter-charging times, an Erlang-2 model may represent larger between-individual variance than the exponential model. Therefore, in general, the shape parameter of Erlang-2 model is less than that of the exponential model. This is more pronounced in the Mixed model, which indicates the corresponding amount of heterogeneity is captured in the model.

Thirdly, heterogeneity in charging regularity is taken into consideration by introducing a mixture of two parametric distributions. It is shown that the inclusion of heterogeneity in charging regularity significantly improves model fit, as is evident from the lowest BIC (Bayesian Information Criterion) value. As we noticed in Eq. (15), the posterior probabilities of PEV user’s membership being in the random user segment or in the regular user segment were obtained. The parameter estimates of φ indicate the relative size of the random users segment. The results suggest that about 90% of the entire sample of PEV users are more likely to charge their car at random intervals, while 10% charges regularly.

Lastly, the effects of covariates on inter-charging time are presented. The estimated coefficients of covariates are statistically significant at the 0.001 level and stable across all models. The signs of coefficient estimates are mostly consistent with their priors. Harsh weather conditions would lower PEV users’ charging rates. Thus, we expect the sign of the coefficient to be negative for all variables. A notable exception, however, was observed for Temperature < 0 °C which has a positive coefficient estimate, implying the covariate increases charging rates. The magnitude of the covariate effects can be interpreted as the percentage change in the hazard, \exp(x_ikβ), as represented by Eq. (10). In the mixed model specification with covariates, on a certain day when the mean temperature is lower than 0 °C, a PEV users’ charging rate increases by 5%. This result may indicate this variable reflects not only PEV users’ charging behavior but also factors affecting battery consumption such as parasitic load (e.g., heating) and battery performance at low temperatures. Several empirical tests support the fact that batteries work evidently less efficiently at low temperatures (e.g., Yuksel and Michalek, 2015).

In contrast, Temperature > 20 °C has a larger negative coefficient estimate compared to others, which leads to a 14% decrease in the charging rate on the corresponding day, suggesting a synergistic effect of covariate effects on charging rate.
and good performance of battery at this temperature. Heavy wind speed and Heavy precipitation have high negative effects on inter-charging times of PEV users at public charging stations, and lower the charging rate by 5% and 4%, respectively.

To summarize, the model estimates are consistent with the formulated hypotheses. The model fit statistics suggest that the model with heterogeneity in charging frequency as well as charging regularity performs the best, and inclusion of time-varying variable significantly improves model fit.

5.2. Characteristic differences between random users and regular users

To get a better understanding of the reasons why people behave differently in terms of charging regularity, we take a look at the characteristics of each segment. Two sets of variables were considered as shown in Table 4: charging behavior and vehicle characteristics. Although demographic characteristics may also explain why heterogeneity in charging regularity arises, their effects could not be estimated because socio-demographic variables were not available due to confidentiality of the data.

We considered five charging behavior variables and two vehicle variables. Charging behavior characteristics included the average charging interval, charging duration, and charging frequency and the percentage of weekend charging. Moreover, we considered charging station loyalty of each user, which was calculated as \( \sum_{i=1}^{N} d_i^2 \), where \( d_i \) is the share of the charging station

---

### Table 2

Description of variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Mean</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heavy wind speed</td>
<td>1 if daily mean wind speed is greater than 8.7 m/s, 0 otherwise</td>
<td>0.062</td>
<td>0.241</td>
</tr>
<tr>
<td>Temperature &lt; 0°C</td>
<td>1 if daily mean temperature is less than 0°C, 0 otherwise</td>
<td>0.060</td>
<td>0.237</td>
</tr>
<tr>
<td>Temperature &gt; 20°C</td>
<td>1 if daily mean temperature is greater than 20°C, 0 otherwise</td>
<td>0.046</td>
<td>0.211</td>
</tr>
<tr>
<td>Heavy precipitation</td>
<td>1 if daily precipitation amount is greater than 11.1 mm, 0 otherwise</td>
<td>0.045</td>
<td>0.207</td>
</tr>
</tbody>
</table>

### Table 3

Model estimation results.

<table>
<thead>
<tr>
<th></th>
<th>Exponential</th>
<th>Erlang-2</th>
<th>Exponential/gamma</th>
<th>Erlang-2/gamma</th>
<th>Mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda )</td>
<td>0.356***</td>
<td>0.718***</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(711.00)</td>
<td>(287.20)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixing distribution parameter: Exponential model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shape (( \gamma_{Exp} ))</td>
<td>–</td>
<td>–</td>
<td>1.916***</td>
<td>–</td>
<td>2.155***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(65.31)</td>
<td></td>
<td>(57.77)</td>
</tr>
<tr>
<td>Scale (( \alpha_{Exp} ))</td>
<td>–</td>
<td>–</td>
<td>6.491***</td>
<td>–</td>
<td>8.116***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(57.00)</td>
<td></td>
<td>(50.19)</td>
</tr>
<tr>
<td>Mixing distribution parameter: Erlang-2 model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shape (( \gamma_{Er} ))</td>
<td>–</td>
<td>–</td>
<td>1.805***</td>
<td>1.800***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(68.02)</td>
<td>(18.42)</td>
</tr>
<tr>
<td>Scale (( \alpha_{Er} ))</td>
<td>–</td>
<td>–</td>
<td>2.984***</td>
<td>1.842***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(58.55)</td>
<td>(19.68)</td>
</tr>
<tr>
<td>Size of exponential (( \phi ))</td>
<td>–</td>
<td></td>
<td></td>
<td>0.900*** (37.18)</td>
<td></td>
</tr>
<tr>
<td>Weather characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heavy wind speed</td>
<td>–0.023**</td>
<td>–0.019**</td>
<td>–0.051**</td>
<td>–0.050***</td>
<td>–0.052***</td>
</tr>
<tr>
<td></td>
<td>(–0.563)</td>
<td>(–0.590)</td>
<td>(–0.617)</td>
<td>(–0.763)</td>
<td>(–0.765)</td>
</tr>
<tr>
<td>Temperature &lt; 0°C</td>
<td>0.008</td>
<td>0.011</td>
<td>0.050**</td>
<td>0.056</td>
<td>0.050**</td>
</tr>
<tr>
<td></td>
<td>(2.22)</td>
<td>(3.85)</td>
<td>(5.80)</td>
<td>(8.86)</td>
<td>(6.24)</td>
</tr>
<tr>
<td>Temperature &gt; 20°C</td>
<td>–0.134***</td>
<td>–0.160***</td>
<td>–0.136**</td>
<td>–0.157***</td>
<td>–0.146***</td>
</tr>
<tr>
<td></td>
<td>(–11.26)</td>
<td>(–17.31)</td>
<td>(–16.33)</td>
<td>(–24.39)</td>
<td>(–18.64)</td>
</tr>
<tr>
<td>Heavy precipitation</td>
<td>–0.067**</td>
<td>–0.086**</td>
<td>–0.040**</td>
<td>–0.048**</td>
<td>–0.042**</td>
</tr>
<tr>
<td></td>
<td>(–4.83)</td>
<td>(–7.95)</td>
<td>(–5.61)</td>
<td>(–8.18)</td>
<td>(–6.18)</td>
</tr>
<tr>
<td>Log-likelihood at convergence (LLc)</td>
<td>–935,259</td>
<td>–1,134,588</td>
<td>–834,454</td>
<td>–919,936</td>
<td>–827,864</td>
</tr>
<tr>
<td>Log-likelihood at zero (LL0)</td>
<td>–1,353,268</td>
<td>–1,353,268</td>
<td>–1,353,268</td>
<td>–1,353,268</td>
<td>–1,353,268</td>
</tr>
<tr>
<td>Number of parameters (N)</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Total number of individuals</td>
<td>9027</td>
<td>9027</td>
<td>9027</td>
<td>9027</td>
<td>9027</td>
</tr>
<tr>
<td>Total number of observations</td>
<td>449,844</td>
<td>449,844</td>
<td>449,844</td>
<td>449,844</td>
<td>449,844</td>
</tr>
<tr>
<td>Adj. likelihood ratio index = 1 – ([LLc – N(LL0)]/LL0)</td>
<td>0.309</td>
<td>0.162</td>
<td>0.383</td>
<td>0.320</td>
<td>0.388</td>
</tr>
<tr>
<td>BIC</td>
<td>1,870,595</td>
<td>2,269,254</td>
<td>1,668,986</td>
<td>1,839,949</td>
<td>1,655,845</td>
</tr>
</tbody>
</table>

Notes: t-values are in the parentheses.

BIC = \( – 2 \cdot LL + N \cdot \log(\text{number of observations}) \). The model with the lowest BIC is preferred. For unobserved heterogeneity concerns, it is assumed that inter-charging time hazard (\( \lambda_i \)) are to be distributed as gamma with a shape parameter and a scale parameter (Mean = \( \gamma/\alpha \), Variance = \( \gamma/\alpha^2 \)).

* Statistical significance at the 0.001 level.
** Statistical significance at the 0.01 level.
*** Statistical significance at the 0.05 level.
In the total number of \( N \) visited charging stations. Vehicle characteristics considered dummy variables for two EV types (BEV and PHEV).

Based on the parameter estimates of the mixed model, 8121 PEV users were assigned to the random group while 906 were assigned to the regular group. Table 5 provides the mean values, the standard deviations, and the corresponding \( t \)-values for mean differences between two groups. All selected variables are statistically different at the 0.001 level. As for the charging behavior characteristics, on average, regular users charge at public charging station almost every 2.75 days, while random users charge every 5.65 days. Also, it is interesting to note that PEV users who have regular charging patterns tend to charge more and longer compared to random users. Weekend charging is more significant in the random user group. That is to say, regular users tend to charge more at weekdays. Moreover, regular users exhibit stronger charging station loyalty than random users. These results are in line with our expectations.

In conclusion, the results indicate that regular users in our data show distinct characteristics from random users for both charging behavior and vehicle variables. In terms of vehicle characteristics, compared with the random group, the regular group is more represented by PHEV owners. This result is somewhat surprising in that one would expect that BEV users charge more regularly as they are more dependent on charging. In the Dutch context, however, the result can be explained by the fact that more than 85% of the PEV market share is dominated by PHEV. Thus, these users dominate the use of public charging stations. Moreover, BEV users mainly charge at home in private parking places or at work (e.g., Accenture et al., 2015).

## 6. Discussion and conclusions

This paper examined PEV users’ inter-charging times to understand their charging patterns at public charging infrastructure with a Dutch four-year longitudinal charging transactions data set. Given that PEV users exhibit heterogeneous charging behavior in terms of charging frequency and charging regularity, this study applied a hazard-based duration model to examine how often PEV users charge their car, and to distinguish the distinct characteristics of random (erratic) users and regular (routine) users.

The results show that 90% of PEV users charge their car randomly at public charging stations, while 10% charge regularly. It is found that significant differences exist between the two groups in terms of charging characteristics, such as charging interval, charging duration, loyalty to charging station and total number of charging episodes. Given the estimated parameters of the model, weather conditions appear to have a direct impact on charging decisions at public charging stations.

Our empirical results have implications for practice and research. For service providers, knowing the different characteristics of random and regular users would be managerially meaningful. Most PEV users tend to charge randomly at public.
charging stations, and they are likely to be light users given the short charging durations and long charging intervals. Regular users, however, those who exhibit habitual process in terms of charging interval, have a high loyalty to a specific charging station.

Moreover, the results can be associated with quantifiable information about the extent weather conditions affect the periodicity of everyday charging activity. Our findings suggest that harsh weather conditions postpone one's decision to charge a PEV at a public charging station. When aiming to stimulate the use of public charging infrastructure, it may be important to bear in mind that reducing weather exposure to PEV users, by means of reduced walking distance or increased shelter, is not a trivial factor.

In terms of modeling approach, recognizing heterogeneity in charging regularity and taking it into account may play a crucial role in successful implementation of future charging infrastructure. Such information is relevant to energy providers and allows them to better manage smart grids, which may provide the input to decide on the optimal location patterns of charging stations. It would be of interest in future research to incorporate models of dynamic charging behavior into computational process models of activity-travel behavior or into scheduling tools to derive essential information concerning the temporal allocation of recurrent behavioral activity-travel patterns. Both researchers and practitioners in the area of management of public charging infrastructure can gain insights to understanding the periodic and repetitive nature of charging behavior.

An interesting future research topic is to examine the effects of PEV users’ socio-demographics on inter-charging time hazard. In this study, we endogenously segmented users into two groups only by considering the patterns of charging intervals. Nevertheless, the models developed in this study provide a basic insight into how PEV users utilize public charging infrastructure. These models can be quickly adapted with more abundant data sources. Given that the charging activity is chained with daily activity schedules and the state-of-charge (SOC), it would be a valuable to examine in future research how those variables affect charging regularity and charging frequency.

Acknowledgement

The authors would like to thank Auke Hoekstra at the Stichting E-laad for making the charging transaction data available for this research. We would also like to thank reviewers for their insightful comments on the paper as these led us to an improvement of the work.

References
