A Comparison of Nonlinear Pricing Preference Models for Digital Services

Completed Research Paper

Christian Schlereth
Goethe-University Frankfurt
Faculty of Business and Economics, Department of Marketing
Grueneburgplatz 1, 60323 Frankfurt am Main, Germany
schlereth@wiwi.uni-frankfurt.de

Abstract

We review and empirically compare a variety of hierarchical Bayes models that use the data obtained from discrete choice experiments to capture usage behavior under two-part tariffs, as are commonly used in service industries such as telecommunication, cloud computing, and online music downloads. The models enable the simultaneous prediction of consumers' service purchase-, tariff choice-, and usage quantity-decisions as well as service provider's profit. We outline recent extensions, which differ with respect to the concepts applied to capture usage uncertainty, estimation space, and complexity of the parameter estimation. We illustrate and compare the models using data from an empirical study about Internet access services. The results show that willingness-to-pay differs significantly across quantity units, a simple estimation procedure leads to inferior results, and usage uncertainty as well as estimation space have only limited influence on the results. We also illustrate how to use these results to determine optimal tariffs.

Keywords: Nonlinear Pricing, discrete choice models, digital services, hierarchical Bayes
Introduction

Two-part tariffs belong to the class of nonlinear pricing plans and commonly appear in service industries such as telecommunications, cloud computing services, banking, online music downloads, and health maintenance (see Iyengar and Gupta 2009 for a review). They frequently consist of a usage-independent fixed (access) fee, paid to gain access to the service, and a marginal price, charged for each unit consumed (Huang and Sundararajan 2011). For example, Amazon charges for a Medium EC2 Standard Reserved Instance a (yearly) fixed fee of €122 and a marginal price of $0.085 per hour. Other service providers use different nonlinear pricing plans, such as three-part tariffs (Masuda and Whang 2006), which also contain an allowance, up to which the consumer can use the service for free and the marginal price is charged for each unit after exceeding the allowance. Other prominent types of tariffs are bucket pricing plans (Schlereth and Skiera 2012), which are a special case of three-part tariffs with no marginal price, flat-rates (Koehler et al. 2010), and fair-use flat-rates (Fritz et al. 2011), which throttles the service quality, after reaching a certain threshold.

Unfortunately, modeling consumers’ choice and usage behavior toward such two-part tariffs and other nonlinear pricing plans is very challenging, because of the interdependency between prices and consumption (Iyengar et al. 2008; Lambrecht et al. 2007). The following example illustrates this interdependency: Assume a consumer who can choose between two tariffs for a digital service such as video-on-demand, one of which features a two-part tariff with a rather high fixed fee and a rather low marginal price (a flat rate is a special case, because it has a marginal price of zero), and another a two-part tariff with a rather low fixed fee and a rather high marginal price (a pay-per-use tariff is a special case, because it has a fixed fee of zero). Both two-part tariffs might be attractive for this consumer, but her or his usage certainly will differ for them, because marginal prices affect usage. The low marginal price for the first two-part tariff likely prompts this consumer to use the service much more than she or he would with the second two-part tariff. These differences in usage and prices affect the attractiveness and profitability of the tariffs, so that their proper evaluation requires methods that reflect the two-way dependence between prices and consumption.

A willingness-to-pay function can account for this interdependency, because it describes the amount that a consumer is willing to pay for a given quantity of a product, which in turn offers a means to account for a different willingness to pay for each quantity unit. Economics and Marketing provide a rich literature that uses transactional data to estimate preferences for nonlinear pricing plans (see Iyengar and Gupta 2009 for a review). This stream of literature emphasizes the importance of accounting for usage uncertainty; however, the reported studies incorporate usage shocks in an inconsistent way. Moreover, transaction data are unavailable for companies that enter new markets or offer new products that have not yet been sold in real market conditions (Wertenbroch and Skiera 2002). Also, despite the high external validity of transaction data, prices may vary minimally and only purchases are observed – and not decisions not to purchase – such that the estimation of willingness-to-pay is often impossible. The use of survey data thus offers assistance in the frequently occurring situations of setting the prices of new digital services and provides the motivation for this study’s focus on survey data.

Most methods that use survey data to estimate willingness-to-pay (WTP) currently apply to single-unit products, which typically include durable goods, such as washing machines, notebook computers, and cars (e.g., Jedidi and Zhang 2002; Wertenbroch and Skiera 2002). These methods, however, are not applicable to situations in which consumers purchase a varying number of quantity units and value the first units more than subsequent units, as is typical for many digital services. Other studies from Lahiri et al. (2013) and Koehler et al. (2010) compare consumers preferences for different types of tariffs, however, the authors neglect concrete prices and the interdependency between prices and consumption.

The aim of this article is to review and empirically compare a variety of models that use survey data with a focus on discrete choice experiments (Louviere et al. 2000) to capture usage behavior in response to two-part tariffs. These models enable the simultaneous prediction of the consumers’ service purchase decisions, their tariff choice decisions, and their usage quantity decisions, which allows for evaluations of the expected revenues and profits. Even though we primarily model tariff choice and consumption under two-part tariffs, we consider our model as generalizable to other types of tariffs with usage dependent components, such as three-part pricing plans, bucket pricing plans, and fair-use flat-rates. This model serves as the root to present and discuss various extensions to the model, which have been prominently
discussed in the literature. The resulting models differ in the estimation space (i.e., preference space versus WTP space), the consideration of usage uncertainty, and the complexity of parameter estimation (one- versus two-step estimation procedure), which raises the question of whether these differences matter and which of the models performs best.

The reminder of this article is organized as follows: First, we present our theoretical background and develop several discrete choice models that use discrete choice data to capture usage behavior in response to two-part tariffs. We then elaborate on different estimation spaces, different ways to capture uncertainty, and different estimation procedures. Next, we describe our empirical study, followed by the empirical results. We illustrate how to use the results to determine optimal tariffs, and finally, we provide concluding remarks.

Theoretical Background

Utility Function

We assume that a consumer \( i \) does not choose more than one tariff \( j \), and we allow the utility of the consumer to vary across periods, such that consumer \( i \) might pick different tariffs in different periods. Furthermore, we assume that consumer \( i \) has a budget \( Y_{i,t} \) in period \( t \) that she or he can spend on \( q_{ij,t} \) units for a service under tariff \( j \) or on \( z_{ij,t} \) units of a composite (outside) good, not observed by the modellers. Let \( U_{i,j}(q_{ij,t}) \) denote the utility consumer \( i \) derives from this consumption of \( q_{ij,t} \) units and \( q_{ij,t} \) denote the utility of consuming \( z_{ij,t} \) units of the composite (outside) good, where \( q_{ij,t} \) indicates the price parameter (frequently called the income parameter in economics). The following direct utility function expresses the utility of consuming \( q_{ij,t} \) units of a service and \( z_{ij,t} \) units of a composite (i.e., not observable outside) good:

\[
U_{i,j}(q_{ij,t}, z_{ij,t}) = U_{i,j,t}(q_{ij,t}) + \sigma_{ij,t} \cdot z_{ij,t} \quad (i \in I, j \in J, t \in T).
\]

(1)

Consumer \( i \) does not have an infinite budget, and his decisions are subject to a budget constraint. Under 2-part tariffs, he pays a billing rate \( R_{i,j,t} \), which can be decomposed into a fixed fee \( F_j \) and a marginal price \( p_j \), such that \( R_{i,j,t} = q_{ij,t} \cdot p_j + F_j \). We assume that decisions about the quantity \( q_{ij,t} \) are not made independently from the prices in the 2-part tariffs, but depend on the marginal price \( p_j \) in tariff \( j \) (we later show why it does not depend on the fixed fee \( F_j \)). Since we don’t know anything about the price per unit of the unobservable composite good, we normalize it to 1 without loss of generality, such that we can write the budget constraint as:

\[
Y_{i,t} \geq R_{i,j,t} + z_{ij,t} = q_{ij,t} \cdot p_j + F_j + z_{ij,t} \quad (i \in I, j \in J, t \in T).
\]

(2)

A utility-maximizing consumer exhausts the budget, so that Equation (2) will be satisfied as an equality. Rearranging and substituting Equation (2) into Equation (1) then leads to the indirect utility function:

\[
U_{i,j,t}(p_j, F_j) = U_{i,j,t}(q_{ij,t}) + \sigma_{ij,t} \cdot (Y_{i,t} - q_{ij,t} \cdot p_j - F_j) \quad (i \in I, j \in J, t \in T).
\]

(3)

We use \( j = 0 \) to describe the situation in which the consumer decides not to use the service and spend the money on something non observable. The corresponding utility is then \( U_{i,0,t} = \varpi_{i,t} Y_{i,t} \).

Consideration of “Unobserved Behavior” in Tariff Choice Decisions

Discrete choice experiments (Louviere et al. 2000), allow for the estimation of parameters of the indirect utility function in Equation (3). Respondents in a survey have to decide in several choice-sets between various hypothetical service offers and a no-purchase option. These service offers may be described by the prices of the two-part tariffs, the quality of the service, and other service characteristics. Each consumer chooses either one of the services or the no-purchase option; we assume that consumer \( i \) chooses from each choice-set \( s \) the alternative (either the most preferred service charged by a two-part tariff or the no-
purchase option) that yields the highest utility, subject to meeting the budget constraint. To account for any additional unobserved factors \( \varepsilon_{i,j,s,t} \) (usually labeled error terms; e.g., Train 2009), we treat them as random and make probability statements \( \Pr_{i,j,s,t} \) about consumers' choice decisions, such that \( \Pr_{i,j,s,t} \) denotes the probability of consumer \( i \) picking tariff \( j \) in choice-set \( s \) and period \( t \):

\[
(4) \quad \Pr_{i,j,s,t} = \Pr(U_{i,j,s} + \varepsilon_{i,j,s,t} > U_{i,j',s} + \varepsilon_{i,j',s,t}; \forall j \neq j') \quad (i \in I, j' \in J_s, s \in S, t \in T).
\]

We assume the error term \( \varepsilon_{i,j,s,t} \) to be independently distributed extreme value (also called type I extreme value, or Gumbel). In turn, \( \theta_{i,t} \) denotes the vector of parameters included in the utility function, and \( d_{i,j,s,t} \) represents the indicator variables that equal 1 if consumer \( i \) chooses tariff \( j \) from choice-set \( s \) and 0 otherwise. The following likelihood function of the mixed logit model enables us to estimate the vector of the parameters \( \theta_{i,t} \):²

\[
(5) \quad L(\Theta) = \prod_{i=1}^{I} \prod_{s=1}^{S} \prod_{j=1}^{J_s} \frac{\exp(U_{i,j,s})}{\exp(U_{i,j,s}) + \sum_{j' \neq j} \exp(U_{i,j',s})} (t \in T).
\]

Note that the income term \( d_{i,s,t} \cdot Y_{i,t} \) gets cancelled out in the likelihood function, because it has no influence on the differences between the utilities of all alternatives.

**Estimation Space**

Equation (3) describes the utility function in preference space (see Train and Weeks 2005) meaning that parameters are estimated as preferences and are subsequently transformed into willingness-to-pay (WTP). In line with Moorthy et al. (1997), we define WTP as the amount of money for which a consumer is indifferent between purchasing and not purchasing:

\[
(6) \quad U_{i,j,t}(q_{i,j,t} + \sigma_{i,t} \cdot (Y_{i,t} - q_{i,j,t} \cdot p_j - F_j)) = \sigma_{i,t} \cdot Y_{i,t} \quad (i \in I, j \in J, t \in T).
\]

Substituting the billing rate (i.e., \( F_j + q_{i,j,t} \cdot p_j \)) in Equation (6) with the WTP for \( q_{i,j,t} \) units, \( WTP_{i,j,t}(q_{i,j,t}) \), and rearranging, we obtain \( WTP_{i,j,t}(q_{i,j,t}) = U_{i,j,t}(q_{i,j,t}) / \sigma_{i,t} \), respectively \( U_{i,j,t}(q_{i,j,t}) = WTP_{i,j,t}(q_{i,j,t}) \cdot \sigma_{i,t} \), which we substitute in Equation (3).

\[
(7) \quad U_{i,j,t}(p_j, F_j) = \sigma_{i,t} \cdot WTP_{i,j,t}(q_{i,j,t}) + \sigma_{i,t} \cdot (Y_{i,t} - q_{i,j,t} \cdot p_j - F_j) \quad (i \in I, j \in J, t \in T).
\]

The indirect utility function in Equation (7) describes an alternative utility function labelled as WTP space (Train and Weeks 2005) or surplus model (Sonnier et al. 2007). The reason is that consumer surplus \( CS_{i,j,t}(p_j, F_j) \) of consumer \( i \) under tariff \( j \) is defined as the difference between the willingness-to-pay \( WTP_{i,j,t}(q_{i,j,t}) \) and the associated billing rate charged for consuming \( q_{i,j,t} \) units, that is:

\[
(8) \quad CS_{i,j,t}(p_j, F_j) = WTP_{i,j,t}(q_{i,j,t}) - q_{i,j,t} \cdot p_j - F_j \quad (i \in I, j \in J, t \in T).
\]

By substituting Equation (8) into Equation (7), we obtain Equation (9), which outlines how consumer surplus relates to the indirect utility function:

\[
(9) \quad U_{i,j,t}(p_j, F_j) = \sigma_{i,t} \cdot CS_{i,j,t}(F_j, p_j) + \sigma_{i,t} \cdot Y_{i,t} \quad (i \in I, j \in J, t \in T).
\]

The advantage of the estimation in WTP space is that the subsequent transformation to WTP is no longer necessary, because all parameters are estimated as monetary values. All indirect utility functions — that is, Equations (3), (7), and (9) — are behaviorally equivalent, and their use in the likelihood function of Equation (5) theoretically should lead to the same results. However, Train and Weeks (2005) and Sonnier et al. (2007) warn that many estimation methods (especially hierarchical Bayes methods) also depend on the common prior distributions (or prior beliefs), and the most common prior distributions usually are not equivalent among these utility functions. Thus, the use of the common prior distributions in hierarchical Bayes methods frequently leads to different parameters for the utility functions. In particular, they show that estimation in the WTP space yields more face-valid parameters and that the WTP derived by the estimated parameters in preference space becomes unreasonably high. In an empirical study by Sonnier

---

² See Train (2009) for a more elaborate discussion of the underlying assumptions.
et al. (2007), the authors show that the unrealistic large WTPs estimated from the preference space eventually result in unrealistic predictions of optimal prices and predicted market shares. We use their considerations about the estimation space for the first time for multi-unit products to estimate the parameters of willingness-to-pay functions.

**Modelling Willingness-to-Pay Functions and Demand Functions**

**Direct Utility Functions**

Subsequently, we extend the model in Equation (3) to explicitly account for two-part tariffs. To specify the functional form of $U_{i,j}(q_{i,j})$, we assume that the utility increases with the quantity being consumed, $dU_{i,j}(q_{i,j})/dq_{i,j} > 0$, but the corresponding marginal utility of each additional unit decreases, such that $d^2U_{i,j}(q_{i,j})/d^2q_{i,j} < 0$ (e.g., Kridel et al. 1993; Maskin and Riley 1984; Oi 1971). We further assume that no income effects, network externalities, tariff-specific preferences, or differentiations of prices across time zones or regions exist. We neglect the differences of other attributes across tariffs; their inclusion already has been described elsewhere (e.g., Schlereth and Skiera 2012).

For the direct utility function $U_{i,j}(q_{i,j})$, we choose the commonly used quadratic functional form with the desired characteristics (e.g., Iyengar et al. 2008; Lambrecht et al. 2007). In addition, we bind it by a maximum utility level and include a saturation level of usage. This maximum utility level guarantees that the quadratic functional form does not decrease after the saturation level, such that $dU_{i,j}(q_{i,j})/dq_{i,j} > 0$ is satisfied. Therefore,

$$U_{i,j}(q_{i,j}) = \begin{cases} 
-a'_{i,j} \cdot q_{i,j} - \frac{b'_{i,j}}{2} \cdot q_{i,j}^2 + c'_{i,j}, & \text{if } q_{i,j} \leq \frac{a'_{i,j}}{b'_{i,j}} \\
\frac{a'_{i,j}}{b'_{i,j}} + c'_{i,j}, & \text{if } q_{i,j} > \frac{a'_{i,j}}{b'_{i,j}} 
\end{cases} \quad (i \in I, j \in J, t \in T).$$

Equation (10) describes the direct utility function in preference space. The parameters $a'_{i,t}$, $b'_{i,t}$, and $c'_{i,t}$ are individual specific; their apostrophes indicate that the parameters relate to the utility function. The corresponding direct utility function in WTP space can be derived easily by substituting Equation (10) into Equation (6):

$$WTP_{i,j}(q_{i,j}) = \begin{cases} 
a_{i,j} \cdot q_{i,j} - \frac{b_{i,j}}{2} \cdot q_{i,j}^2 + c_{i,j}, & \text{if } q_{i,j} \leq \frac{a_{i,j}}{b_{i,j}} \\
\frac{a_{i,j}}{b_{i,j}} + c_{i,j}, & \text{if } q_{i,j} > \frac{a_{i,j}}{b_{i,j}} 
\end{cases} \quad (i \in I, j \in J, t \in T),$$

where $a_{i,j} = a'_{i,j} / \sigma_{i,j}; b_{i,j} = b'_{i,j} / \sigma_{i,j};$ and $c_{i,j} = c'_{i,j} / \sigma_{i,j}$.

We denote the direct utility function in WTP space in Equation (11) as the willingness-to-pay function, because it describes the amount that a consumer is willing to pay for a given quantity of a service. The parameter $c_{i,t}$ addresses a WTP for a quantity of zero, which is common if the service provider does not charge a marginal price for additional services that a consumer might use within a contract; for example consumers in Europe or Australia, who have subscribed to a cellular phone service contract, do not have to pay for incoming calls. They only pay a marginal price for outgoing calls. The marginal willingness-to-pay function ($MWTP_{i,j}(q_{i,j})$) is the derivative of Equation (11), which indicates the amount consumer $i$ is willing to pay for the $q^{th}$ unit increment:

$$MWTP_{i,j}(q_{i,j}) = \begin{cases} 
a_{i,j} - b_{i,j} \cdot q_{i,j}, & \text{if } q_{i,j} \leq \frac{a_{i,j}}{b_{i,j}} \\
0, & \text{if } q_{i,j} > \frac{a_{i,j}}{b_{i,j}} 
\end{cases} \quad (i \in I, j \in J, t \in T).$$
The demand function \( q_{i,j,t}(p_j) \) is the inverse of the \( i \)th consumer’s marginal willingness-to-pay function from Equation (12) if we substitute for marginal willingness to pay with the marginal price \( p_j \) of the tariff \( j \):

\[
q_{i,j,t} = \begin{cases} 
\frac{a_{i,t} - 1}{b_{i,t}} \cdot p_j & \text{if } p_j \leq a_{i,t} \\
0 & \text{if } p_j > a_{i,t}
\end{cases} \tag{13}
\]

(i \( \in \mathbb{I}, j \in \mathbb{J}, t \in \mathbb{T} \)),

Equation (13) shows that decisions about the quantity \( q_{i,j,t} \) can be inferred by the marginal price \( p_j \) of the 2-part tariffs. It is that quantity, which maximizes the difference between willingness-to-pay and the price to be paid for a certain number of units. Note that the quantity does not depend on the fixed fee. Equation (13) requires that \( p_j \leq a_{i,t} \), which ensures non-negative quantities. In contrast, Equation (12) requires that \( q_{i,j,t} \leq a_{i,t}/b_{i,t} \), to guarantee a non-negative marginal WTP for each unit increment.

### Indirect Utility Functions

Substituting Equation (10) into Equation (3) enables us to derive the indirect utility function (here expressed in preference space):

\[
U_{i,j,t}(p_j,F_j) = \left( \frac{(a'_{i,t} - \sigma_{i,t} \cdot p_j)^2}{2 \cdot b'_{i,t}} + c'_{i,t} + \sigma_{i,t} \cdot (Y_{i,t} - F_j) \right)^\frac{1}{2}, \quad \text{if } p_j \leq a'_{i,t} \tag{14}
\]

\[
\sigma_{i,t} \cdot c'_{i,t} + \sigma_{i,t} \cdot (Y_{i,t} - F_j), \quad \text{if } p_j > a'_{i,t} \]

Then, we can determine consumer surplus \( CS_{i,j,t}(p_j,F_j) \) by inserting Equation (14) into Equation (9):

\[
CS_{i,j,t}(p_j,F_j) = \frac{1}{\sigma_{i,t}} \cdot \left( \frac{(a'_{i,t} - \sigma_{i,t} \cdot p_j)^2}{2 \cdot b'_{i,t}} + c'_{i,t} \right)^\frac{1}{2} - F_j \tag{15}
\]

\[
\frac{1}{\sigma_{i,t}} \cdot c'_{i,t} - F_j \quad \text{if } p_j > a'_{i,t}
\]

In Table 1, we summarize these considerations for how to specify the indirect utility function and determine WTP for \( q_{i,j,t} \) units, consumer surplus, the saturation level, and the maximum WTP. The Equations in Table 1 distinguish between both estimation spaces, that is, the preference space – i.e., the estimation of \( a'_{i,t}, b'_{i,t}, \) and \( c'_{i,t} \) and WTP transformation after estimation - and the WTP space – i.e., the direct monetary estimation, by replacing \( a'_{i,t}, b'_{i,t}, \) and \( c'_{i,t} \) through \( a'_{i,t}, \sigma_{i,t}, b_{i,t}, \sigma_{i,t}, \) and \( c_{i,t}, \sigma_{i,t} \).

### Table 1: Utility, Willingness-to-Pay, and Demand Functions in Preference Space and WTP Space

<table>
<thead>
<tr>
<th>Function Type</th>
<th>Preference Space Formula</th>
<th>WTP Space Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indirect utility function</td>
<td>( U_{i,j,t}(p_j,F_j) = \frac{1}{\sigma_{i,t}} \cdot \left( \frac{(a'<em>{i,t} - \sigma</em>{i,t} \cdot p_j)^2}{2 \cdot b'<em>{i,t}} + c'</em>{i,t} \right) + \sigma_{i,t} \cdot (Y_{i,t} - F_j) )</td>
<td>( U_{i,j,t}(p_j,F_j) = \sigma_{i,t} \cdot \left( \frac{(a'<em>{i,t} - p_j)^2}{2 \cdot b</em>{i,t}} + c_{i,t} \right) + \sigma_{i,t} \cdot (Y_{i,t} - F_j) )</td>
</tr>
<tr>
<td>Willingness-to-pay function</td>
<td>( WTP_{i,j,t}(q_{i,j,t}) = \frac{1}{\sigma_{i,t}} \cdot \left( \frac{(a'<em>{i,t} - \sigma</em>{i,t} \cdot p_j)^2}{2 \cdot b'<em>{i,t}} + c'</em>{i,t} \right) + \sigma_{i,t} \cdot (Y_{i,t} - F_j) )</td>
<td>( WTP_{i,j,t}(q_{i,j,t}) = \frac{1}{\sigma_{i,t}} \cdot \left( \frac{(a'<em>{i,t} - p_j)^2}{2 \cdot b</em>{i,t}} + c_{i,t} \right) + \sigma_{i,t} \cdot (Y_{i,t} - F_j) )</td>
</tr>
<tr>
<td>Consumer surplus</td>
<td>( CS_{i,j,t}(p_j,F_j) = \frac{1}{\sigma_{i,t}} \cdot \left( \frac{(a'<em>{i,t} - \sigma</em>{i,t} \cdot p_j)^2}{2 \cdot b'<em>{i,t}} + c'</em>{i,t} \right) - F_j )</td>
<td>( CS_{i,j,t}(p_j,F_j) = \frac{1}{\sigma_{i,t}} \cdot \left( \frac{(a'<em>{i,t} - p_j)^2}{2 \cdot b</em>{i,t}} + c_{i,t} \right) - F_j )</td>
</tr>
<tr>
<td>Demand function</td>
<td>( q_{i,j,t}(p_j) = \frac{a_{i,t}}{b_{i,t}} )</td>
<td>( q_{i,j,t}(p_j) = \frac{a_{i,t}}{b_{i,t}} )</td>
</tr>
<tr>
<td>Saturation level</td>
<td>( \varpi_{i,t} = \frac{a_{i,t}}{b_{i,t}} )</td>
<td>( \varpi_{i,t} = \frac{a_{i,t}}{b_{i,t}} )</td>
</tr>
</tbody>
</table>
| Indicator – test marginal price | \( \text{Ind}_{i,j,t} = \begin{cases} 
0, & \text{if } a_{i,t} \cdot \varpi_{i,t} \leq a_{i,t} \\
1, & \text{if } a_{i,t} \cdot \varpi_{i,t} > a_{i,t}
\end{cases} \) | \( \text{Ind}_{i,j,t} = \begin{cases} 
0, & \text{if } p_j \leq a_{i,t} \\
1, & \text{if } p_j > a_{i,t}
\end{cases} \) |

Notes: \( a'_{i}, a_{i}, c'_{i}, c_{i} \geq 0, b'_{i}, b_{i} > 0 \)
Treatment of Usage Uncertainty

Thus far, we have assumed that consumer $i$ can pick different tariffs in different periods. Yet for most services, consumers must enter into a long-term contract, spanning several periods, without the possibility of switching to a different tariff in each period $t$ (or may switch only with additional costs). When choosing a tariff, the consumer does not know exactly how often she or he will use the service during future periods $t$. Instead, consumer $i$ has some expectations about mean usage and usage variation — also called usage uncertainty — that she or he takes into account during tariff choices. In line with Lambrecht et al. (2007) and Iyengar et al. (2008), among others, we account for usage uncertainty by incorporating a usage shock $v_{i,t}$ that reflects random variations in usage. This usage shock $v_{i,t}$ is assumed to be normally distributed with mean 0 and standard deviation $\delta$. Hence, we assume that when choosing a tariff, the consumer knows the usage shock $v_{i,t}$ only in distribution and does not know the exact parameter values of her or his willingness-to-pay function in a certain future period $t$. Following the tariff choice, but prior to making the actual usage decision, the consumer learns about the usage shock $v_{i,t}$ in each period, which remains unobserved by the modellers.\(^3\)

To capture usage uncertainty, existing studies model the usage shock $v_{i,t}$ differently. Lambrecht et al. (2007) and Narayanan et al. (2007) let the usage shock $v_{i,t}$ influence the intercept of the demand function (here, parameter $a_i$) but not any of the other parameters. In contrast, Iyengar et al. (2008) let the usage shock influence the quantity $q_{i,j}$ but not the parameters of the willingness-to-pay function.

Because none of these studies compares the different models, we elaborate more on their resulting differences. Linking usage shocks to the parameters of the willingness-to-pay functions assumes that the consumer still makes optimal quantity decisions, though they may vary across periods. The typical example would be a consumer who knows at the beginning of a period that she or he will be going on a business trip and will need to use her or his cellular phone more intensively or, alternatively, will be going on holiday, during which her or his cellular phone usage will decline. Both shifts could be captured by a shift (upward or downward) of the intercept of the demand function. In contrast, linking the usage shock to the quantity $q_{i,j}$ assumes that a consumer has the same willingness-to-pay function across all periods but deviates from optimal usage because of stochastic behavior, which is caused by the usage shock $v_{i,t}$.

To illustrate the different effects of these approaches on usage, WTP, billing rate and consumer surplus, we construct in Table 2 a simple numerical example with two considered periods, $t_1$ and $t_2$, during which a consumer can be described by the parameters $a_i = .5$, $b_i = .005$, and $c_i = 1$. We set the fixed fee $F_i$ to 10 Euro and the marginal price $p_j$ to .10 Euro. To simplify our numerical illustration, we consider two instances of the usage shock $v_{i,t}$ that differ only in their sign, such that the mean value of the usage shock $v_{i,t}$ is still distributed symmetrically around 0. The instances of the usage shock $v_{i,t1}$ and $v_{i,t2}$ for parameter $a_i$ take the values .2 and -.2; for parameter $b_i$, they are .001 and -.001; for parameter $c_i$, they equal .5 and -.5; and for the quantity $q_{i,j}$, they are 8 and -8.

The usage shocks influence the quantity, WTP, and billing rate in all models differently. The symmetrical distribution of the usage shock $v_{i,t}$ means that the average quantity and, thus, the average billing rate do not depend on usage shocks $v_{i,t}$ that are related to the parameters $a_i$ or $c_i$ as well as the quantity. The reason is that in case of an influence of the parameter $a_i$ or the quantity, the shock increases or decreases the quantity by the same amount. In contrast, usage shocks on the parameter $c_i$ never affect the quantity. Only usage shocks on the parameter $b_i$ influence the average quantity, because of the quadratic term in the willingness-to-pay function.

Yet usage uncertainty influences average WTP in all cases except the one in which it is linked to the parameter $c_i$. Symmetrical changes of the quantity have a non-symmetrical influence on WTP (see Equation (10)), and changes in the parameter $c_i$ have no effect on the quantity $q_{i,j}$ (see Equation (13)). Notably, usage uncertainty increases consumer surplus and consequently the utility for the consumer if it influences the two parameters $a_i$ and $b_i$ of the willingness-to-pay function, but decreases consumer surplus if it directly influences the usage quantity $q_{i,j}$.

---

\(^3\) Some studies also account for usage uncertainty, noted after tariff choice or through learning (Narayanan et al. 2007). However, capturing such usage uncertainty requires knowledge about the actual consumption, which is usually available in transaction data but unavailable in most survey data.
Table 2: Numerical Example of Different Approaches to Capture Usage Uncertainty

<table>
<thead>
<tr>
<th>Usage Shock $\upsilon_{ij}$ on $b_{ij}$</th>
<th>Usage Shock $\upsilon_{ij}$ on $c_{ij}$</th>
<th>Usage Shock $\upsilon_{ij}$ on $q_{ij}$</th>
<th>Usage Shock $\upsilon_{ij}$ on $a_{ij}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Shock</td>
<td>Shock in $t_1$</td>
<td>Shock in $t_2$</td>
<td>Shock in $t_1$</td>
</tr>
<tr>
<td>$\delta$</td>
<td>$\delta$</td>
<td>$\delta$</td>
<td>$\delta$</td>
</tr>
<tr>
<td>$\tau$</td>
<td>$\tau$</td>
<td>$\tau$</td>
<td>$\tau$</td>
</tr>
<tr>
<td>Quantity</td>
<td>80</td>
<td>120</td>
<td>40</td>
</tr>
<tr>
<td>WTP (in Euro)</td>
<td>25.00</td>
<td>49.00</td>
<td>9.00</td>
</tr>
<tr>
<td>Billing rate (in Euro)</td>
<td>18.00</td>
<td>22.00</td>
<td>14.00</td>
</tr>
<tr>
<td>Consumer surplus (in Euro)</td>
<td>7.00</td>
<td>27.00</td>
<td>-5.00</td>
</tr>
<tr>
<td>Average quantity (in Euro)</td>
<td>80</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>Average WTP (in Euro)</td>
<td>25.00</td>
<td>29.00</td>
<td>26.00</td>
</tr>
<tr>
<td>Average billing rate (in Euro)</td>
<td>18.00</td>
<td>18.00</td>
<td>18.33</td>
</tr>
<tr>
<td>Average consumer surplus (in Euro)</td>
<td>7.00</td>
<td>11.00</td>
<td>7.67</td>
</tr>
</tbody>
</table>

Estimation Procedure

The estimation of the parameters relies on random utility theory and employs the mixed logit model with hierarchical Bayes estimation. Let the vector $\theta_i = (a_i, b_{ij}, c_{ij}, ...)$ summarize the parameters of consumer $i$ in period $t$. Because at the time of the tariff choice, the consumer only knows the parameter values in the form of distributions over all future periods, we can decompose $\theta_i$ into timely invariant parameters $\theta = (a_i, b_i, c_i, ...)$ and the usage shock $\upsilon_{ij}$. The usage shock $\upsilon_{ij}$ is unknown to the consumer, so its instances are derived by drawing from a quasi-random variable, which is normally distributed $N(\alpha, \delta^2)$.

We specify the density of each parameter of $\theta_i$ to be log-normally distributed, with mean $\bar{\theta}$ and covariance matrix $\Omega$, denoted by $g(\theta | \bar{\theta}, \Omega)$. The log-normal distributions ensure positive parameter values, with which we obtain the desired convex functional form of the willingness-to-pay function. The standard deviation $\delta_i$, which specifies the usage shock $\upsilon_{ij}$, also follows a log-normal distribution with mean $\bar{\delta}$ and standard deviation $\tau$, denoted by $f(\delta | \bar{\delta}, \tau)$. Because the usage shock $\upsilon_{ij}$ is stochastic to the consumer prior to the tariff choice, we determine the expected utility or consumer surplus (see Table 1) by integrating over all possible instances of the usage shock $\upsilon_{ij}$. The conditional posterior on $\theta, \delta, \forall i$, given $\bar{\theta}, \bar{\delta}, \Omega,$ and $\tau$, is:

$$
\Lambda(\theta, \delta | \bar{\theta}, \bar{\delta}, \Omega, \tau) \propto \prod_{i \in J} \prod_{s \in S} \prod_{j \in J} \int L(\theta_i, \delta_i, \Sigma) \cdot g(\theta_i | \bar{\theta}, \Omega) f(\delta_i | \bar{\delta}, \tau) d\delta d\theta,
$$

for which the likelihood function is determined according to Equation (5). It can incorporate the in-direct utility function in either the preference space or the WTP space (see Table 1).

The stochastic integration over all instances of the usage shock $\upsilon_{ij}$ is cumbersome to calculate, because adding a quasi-random, normally distributed variable may result in unrealistic instances, such as usage quantities $q_{ij}$ below zero or above the saturation level, as well as negative parameter estimates for $a_i$ or $b_i$. To avoid such instances, truncations or exponentiations usually serve to restrict the limits of the integrals, even though, they effectively lead to unrealistic shock distributions. The required steps differ across approaches to capture usage uncertainty and thus might influence the estimation results. In Table 3, we detail the required steps to capture the usage shock $\upsilon_{ij}$ on $a_i, b_i,$ and $q_{ij}$. We no longer consider the linkage of usage uncertainty with $c_i$, because it has no impact on the usage quantity.

Note that the estimation that does not capture usage uncertainty is a special case of the conditional posterior described in Equation (16). The assumption that the consumer $i$ can always switch between tariffs or unsubscribe in each period $t$ at no additional costs implies that the consumer is certain about her or his usage quantity decision during the actual period, and that she or he has no need to consider future periods, when making tariff choice decisions. Hence, the standard deviation $\delta_i$ may be fixed to 0 for all consumers $i$, which reduces the conditional posterior to $\Lambda(\theta | \bar{\theta}, \Omega) \propto \prod_{i \in J} \prod_{s \in S} \prod_{j \in J} L(\theta_i, \Sigma) \cdot g(\theta_i | \bar{\theta}, \Omega)$.
Linking usage uncertainty with parameter $a_i$ requires no additional truncations, because non-negative instances of $a_i$ are already ensured by the non-negative marginal price. However, linking usage uncertainty with the parameter $b_i$ requires one additional truncation. Because small values of $b_i$ near zero result in very large saturation levels, we restrict the value of $b_i$ to be positive and require $b_i > a_i/q_{\text{max}}$, where $q_{\text{max}}$ is a reasonable constitutional threshold, such as ongoing usage of the service throughout the whole period. The linkage of usage uncertainty to the quantity $q_{ij}$ requires three truncations: The first truncation ensures non-negative usage quantities, the second truncation guarantees non-negative usage quantities $q_{ij}$, and the third truncation ensures that the usage quantity is always below the saturation level. This last truncation guarantees that the assumption of non-decreasing utilities is fulfilled.

### Two-step Estimation Procedure

Although the maximization of the likelihood function in Equation (5) simultaneously estimates the parameters $a_{il}, b_{il}, c_{il}$, and $e_{il}$, it requires a nonstandard utility specification, which is non-additive and therefore not supported by standard estimation software for discrete choice models, such as Limdep, Latent Gold, or Sawtooth Software. We outline a less elaborate estimation procedure for the parameters of the willingness-to-pay functions that basically divides the one-step estimation procedure of the likelihood function in Equation (5) into two steps. The first step consists of a "classical" estimation of a discrete choice model, which leads to parameters that describe the partworth utilities for each marginal price and the price parameter $e_{il}$. Most available software packages for discrete choice models can easily do so. The second step takes the parameters of the first step to derive the parameters $a_{il}$, $b_{il}$, and $c_{il}$ of the willingness-to-pay functions.

Let $K$ be the index set of different marginal prices $p_i$ and $L$ be the index set of different fixed fees $F_i$. A tariff $j$ consists of one of the possible $|K| \cdot |L|$ combinations between a certain marginal price $p_i$ and a certain fixed fee $F_i$. We recommend using a small number of different marginal prices and fixed fees compared with the number of choice-sets, though at least four, to identify the three parameters of the willingness-to-pay functions.
pay function properly. We substitute the index \( j \) with index \( k \) for the different marginal prices, and \( l \) for the different fixed fees, so that we can rewrite Equation (3) as:

\[
U_{i,k,l,t}(p_k, F_i) = U_{i,k,l}(q_{i,k,t}) - \sigma_{i,j} \cdot q_{i,k,t} \cdot p_k + \sigma_{i,j} \cdot (Y_{i,j} - F_i) \quad (i \in I, k \in K, l \in L, t \in T).
\]

The following transformation seeks to capture all nonlinear parts of the indirect utility function through partworth utilities. Therefore, we substitute for any of the \(|K|\) marginal prices \( p_k \) with:

\[
pw_{i,k,t} = U_{i,k,t}(q_{i,k,t}) - \sigma_{i,j} \cdot q_{i,k,t} \cdot p_k \quad (i \in I, k \in K, l \in L, t \in T).
\]

Those partworth utilities \( pw_{i,k,t} \) represent the difference between the utility of consuming \( q_{i,k,t} \) units and the portion of the billing rate that is due to consuming quantity \( q_{i,k,t} \) (expressed in utility terms by multiplying it with the price parameter \( \sigma_{i,j} \)). Therefore, the \(|K|\) partworth utilities summarize all components in the indirect utility function of Equation (17), which depend on the marginal price. This substitution allows us to rewrite the indirect utility function in Equation (17) as an additive function:

\[
U_{i,k,l,t}(pw_{i,k,t}, F_i) = pw_{i,k,t} + \sigma_{i,j} \cdot (Y_{i,j} - F_i) \quad (i \in I, k \in K, l \in L, t \in T).
\]

The use of Equation (19) leads to the likelihood function specification in Equation (20), in which the binary variables \( x_{k,j} \) equal 1 if the marginal price \( p_k \) is present in tariff \( j \) and 0 otherwise:

\[
L | (\theta_j, \Sigma_j) = \prod_{i \in I} \prod_{k \in S} \prod_{j \in J} \left\{ \frac{\exp\left( \sum_{i,k} pw_{i,k,t} \cdot x_{i,j} + \sigma_{i,j} \cdot F_j \right)}{1 + \sum_{j' \neq j} \exp\left( \sum_{i,k} pw_{i,k,t} \cdot x_{i,j'} + \sigma_{i,j} \cdot F_j \right)} \right\} (t \in T).
\]

Most commercially available discrete choice software packages can easily estimate this likelihood function. With Equation (8), we can then calculate the consumer surplus by substituting with Equations (6) and (18) and rearranging:

\[
CS_{i,k,l,t}(pw_{i,k,t}, F_i) = \frac{pw_{i,k,t} - F_i}{\sigma_{i,j}} \quad (i \in I, k \in K, l \in L, t \in T).
\]

The second step builds on the idea that the consumer surplus \( CS_{i,k,l,t}(pw_{i,k,t}, F_i) \) in Equation (21) must be as close as possible to the consumer surplus in Equation (15), which is defined according to the parameters of the willingness-to-pay function. Therefore, we use a Gradient method to minimize Equation (22):

\[
\sum_{i,k} c_{i,k,t}^2 = \sum_{i,k} \left( CS_{i,k,t}^{\text{adj}} - CS_{i,k,t}^{\text{WTP}} \right)^2 \rightarrow \text{Min!} \quad (i \in I, t \in T, a_{t,i}, b_{t,i}, c_{t,i} \geq 0).
\]

The major advantage of this two-step estimation procedure, compared with the preceding one-step estimation procedure, is that various commercially available software packages can estimate the parameters in the first step, and the minimization problem in the second step is relatively easy to solve. This advantage, however, comes at the cost of having two error terms (one for the estimation of the parameters of the partworth utilities and the price in Equation (20), and one for the estimation of the parameters of the willingness-to-pay function in Equation (22)) instead of just one error term (Equation (5)). In addition, estimation in WTP space is not possible, nor is the inclusion of usage uncertainty.

**Empirical Study**

As we outline in the previous section, a variety of models can predict usage behavior in response to two-part tariffs; they differ with respect to how they capture uncertainty, the estimation space, and the complexity of the parameter estimation. The crucial question is which of these models performs best in an empirical setting. Therefore, we use the data from a discrete choice experiment published in Schlereth et al. (2011) that empirically compares the performance of the models with respect to their internal, face, and predictive validity.
Study Design

Respondents were asked in an online survey about their preferences for having access to the Internet. After reporting on their current Internet usage, they chose one alternative among two different tariffs and a no-purchase option from 21 choice-sets (illustrated in Figure 1). All presented tariffs combine monthly fixed fees and marginal prices, ranging between 11 and 32 Euro for the fixed fee and .30 and 1.20 Euro per hour for the marginal price, which results in a 4² design. We never asked about the intended quantity under certain prices of the 2-part tariffs, because respondents are hardly able to do so for many 2-part tariffs. Instead, we inferred the quantity from parameter estimates using Equation (13). The set-up of the study is realistic, because at the time of the study, most tariffs contained usage dependent prices and flat-rates were rather expensive and as such chosen by the minority. The consideration of other attributes is also possible (e.g., Schlereth and Skiera 2012), but we avoid it to focus exclusively on the different impact of prices on usage behavior across the different models.

![Figure 1: Illustration of one of the choice-sets](image)

We use 19 choice-sets to estimate the parameters of the willingness-to-pay functions and evaluate the internal validity; we employ 2 choice-sets to evaluate the predictive validity. To extent the predictive validity analysis, we also ask respondents to make decisions in 5 additional choice-sets, which contain special cases, namely, flat rates (i.e., marginal price set to 0) and pay-per-use tariffs (i.e., fixed fee set to 0), as well as a no-purchase option. These prediction tasks are fairly challenging, because neither flat rates nor pay-per-use tariffs appear as part of the 19 choice-sets used for the estimation.

The online survey, conducted among undergraduate and graduate students of a major German university in 2006, resulted in 206 completed questionnaires. 76% of the respondents stated that they were paying for Internet by themself. The respondents reported an average Internet usage of 54.49 hours per month. They also indicated high average knowledge about their current Internet usage (4.18 on a five-point rating scale).

Overview of Employed Models

We estimate the ten models in Figure 2. The first is a main effects model, in which the marginal price and fixed fee are coded linearly according to \( U_{ij} = \beta_{i,0} + \beta_{i,1} p_j + \beta_{i,2} F_j \). The results for this fairly simple model serve as a benchmark for the more elaborate models (for a similar approach, see Iyengar et al. 2008). Models 2–5 simultaneously estimate all parameters in the preference space; Models 6–9 do so for the WTP space. Models 2 and 6 do not include usage uncertainty. Models 3, 4, and 5 and, respectively, 7, 8 and 9, link usage uncertainty to the parameters \( a_i \), \( b_i \) or the quantity \( q_{i,j} \). Model 10 uses the two-step estimation procedure, which is feasible using standard software.

![Figure 2: Overview of Employed Models](image)

Every model employs a standard diffuse prior distribution for \( \beta \) and \( \delta \) with a mean close to zero (0.1) and a large standard deviation. For \( \Omega \) and \( \tau \), we use an inverted Wishart distribution with low degrees of freedom. The estimation of the parameters relies on random utility theory and employs the mixed logit
model with hierarchical Bayes. The Bayesian algorithm to estimate Models 1–9 is developed in Matlab and adopts the estimation algorithm for additive normally distributed parameters, as described by Train (2009). It also uses the Gibbs sampler to obtain quasi-random draws of the conditional posterior distribution and applies the Metropolis-Hasting algorithm to draw from these conditional posterior distributions. The estimation of Model 10 relies on standard estimation software — in this case, Sawtooth for the first step and the gradient method available in the Microsoft Excel Solver for the second step.

The subsequently reported results for each model are based on 5,000 iterations that we retained after discarding the initial 20,000 iterations and saving every third iteration (\( \hat{\tau} = 35,000 \) iterations in total). We assess convergence according to the trace plot of the likelihood and parameter values. To interpret the results from a purely classical perspective, we summarize the posterior distributions by calculating the mean values of the draws.

**Results of Empirical Study**

**Internal and Predictive Validity.**

To evaluate the internal validity, we calculate the log marginal density (LMD), as well as the log-Bayes factor (Log-BF), the (internal) first choice hit rate for the 19 choice-sets, and the deviance information criterion (DIC). We report the results in Table 4. Model 7, which uses the one-step estimation procedure in WTP space, and which captures the usage uncertainty on parameter \( a_i \), performs best. It attains the highest LMD of -1,354.4 and lowest DIC of 2,737.2. Its Log-BF with respect to the simple benchmark Model 1 is 22.2, which is greater than the critical value of 10 (Kass and Raftery 1995) and supports the higher internal validity of Model 7. Yet, the differences between the (one-step) Models 1–9 are rather small, which indicates that they all have rather comparable internal validity scores.

The first choice (external) hit rates for the two holdouts range between 74.5% for Model 9 and 78.9% for Model 1, such that they are not significantly different from each other \((p > .05)\). Yet the corresponding (external) first choice hit rates for the five holdouts, which include flat rates and pay-per-use tariffs, range between 49.1% for Model 1 and 54.7% for Model 10. They therefore indicate that the simple (benchmark) Model 1 performs worst, whereas the results for the (one-step) Models 2–9 are fairly similar. The two-step model performs best. Not surprisingly, the first choice hit rates are lower for the five than the two holdouts, because the five holdouts contain flat rates and pay-per-use tariffs, which do not appear in the 19 choice-sets that we used to estimate the discrete choice models.

**Face Validity**

To test for face validity, we determine in Table 4 the mean parameter values of the willingness-to-pay functions, the 95% confidence intervals, and the inferred WTP for various quantities. To facilitate the comparison, we divide the reported values of the models estimated in the preference space (Model 2–5) by their price parameter. As expected for Model 1, the values for \( \beta_{i,1} \) and \( \beta_{i,2} \) are negative, which indicates that a lower marginal price and lower fixed fee are preferred to higher levels. The mean values of parameter \( a_i \), which drives the increase of the willingness-to-pay function in the remaining models, range from 1.77 to 2.25. Also as expected, the value of the parameter \( a_i \) is higher than that of parameter \( b_i \), which is responsible for the decrease in the marginal WTP and ranges from .08 to .25. The mean of parameter \( c_i \) describes the usage-independent WTP and varies between 1.73 and 3.11. The price parameter values \( \varpi_i \) range between .41 and .69.

Next, the parameter \( \delta_i \) describes the individual usage uncertainty. Compared with the values of the parameters of \( a_i \) and \( b_i \) and the quantity \( q_{i,j} \), its values are rather small, but they are similar across estimation spaces. Specifically, it ranges between two hours per month for Models 5 and 9 and five hours in Models 4 and 8. This result indicates that average monthly usage uncertainty is small.
### Table 4: Internal and Predictive Validity as well as Description of the Willingness-to-Pay Functions

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal validity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LMD</td>
<td>-1,376.2</td>
<td>-1,373.8</td>
<td>-1,363.9</td>
<td>-1,365.3</td>
<td>-1,373.8</td>
<td>-1,372.1</td>
<td>-1,354.4</td>
<td>-1,360.7</td>
<td>-1,371.7</td>
<td>n.a.</td>
</tr>
<tr>
<td>LogBF</td>
<td>21.8</td>
<td>19.4</td>
<td>9.3</td>
<td>12.1</td>
<td>19.4</td>
<td>17.7</td>
<td>-</td>
<td>6.3</td>
<td>11.7</td>
<td>n.a.</td>
</tr>
<tr>
<td>First choice hit rate</td>
<td>89.2%</td>
<td>88.3%</td>
<td>88.2%</td>
<td>88.3%</td>
<td>88.3%</td>
<td>88.3%</td>
<td>88.3%</td>
<td>88.3%</td>
<td>88.4%</td>
<td>81.3%</td>
</tr>
<tr>
<td>DIC</td>
<td>3,001.4</td>
<td>2,850.0</td>
<td>2,901.1</td>
<td>2,905.6</td>
<td>2,931.4</td>
<td>2,885.6</td>
<td>2,737.2</td>
<td>2,847.3</td>
<td>2,885.0</td>
<td>n.a.</td>
</tr>
<tr>
<td><strong>Predictive validity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First choice hit rate (2 holdouts)</td>
<td>78.9%</td>
<td>75.5%</td>
<td>76.2%</td>
<td>75.7%</td>
<td>76.9%</td>
<td>74.8%</td>
<td>75.0%</td>
<td>74.8%</td>
<td>74.5%</td>
<td>78.6%</td>
</tr>
<tr>
<td>First choice hit rate (5 pay-per-use and flat rate holdouts)</td>
<td>49.1%</td>
<td>49.9%</td>
<td>49.8%</td>
<td>49.9%</td>
<td>50.2%</td>
<td>50.6%</td>
<td>50.6%</td>
<td>50.9%</td>
<td>50.5%</td>
<td>54.7%</td>
</tr>
<tr>
<td><strong>Estimates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a&lt;sub&gt;i&lt;/sub&gt;, β&lt;sub&gt;i,0&lt;/sub&gt;</td>
<td>12.66</td>
<td>2.25</td>
<td>1.94</td>
<td>1.97</td>
<td>1.82</td>
<td>1.91</td>
<td>1.77</td>
<td>1.81</td>
<td>2.00</td>
<td>2.02</td>
</tr>
<tr>
<td>b&lt;sub&gt;i&lt;/sub&gt;, β&lt;sub&gt;i,1&lt;/sub&gt;</td>
<td>9.79</td>
<td>2.25</td>
<td>1.94</td>
<td>1.97</td>
<td>1.82</td>
<td>1.91</td>
<td>1.77</td>
<td>1.81</td>
<td>2.00</td>
<td>2.02</td>
</tr>
<tr>
<td>c&lt;sub&gt;i&lt;/sub&gt;</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>a&lt;sub&gt;0&lt;/sub&gt;, β&lt;sub&gt;0,0&lt;/sub&gt;</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>b&lt;sub&gt;0&lt;/sub&gt;, β&lt;sub&gt;0,1&lt;/sub&gt;</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>a&lt;sub&gt;1&lt;/sub&gt;, β&lt;sub&gt;1,0&lt;/sub&gt;</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>b&lt;sub&gt;1&lt;/sub&gt;, β&lt;sub&gt;1,1&lt;/sub&gt;</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>c&lt;sub&gt;1&lt;/sub&gt;</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>WTP(q=0) (in Euro)</td>
<td>n.a.</td>
<td>1.93</td>
<td>2.71</td>
<td>2.37</td>
<td>7.70</td>
<td>2.29</td>
<td>3.11</td>
<td>3.08</td>
<td>5.71</td>
<td>1.92</td>
</tr>
<tr>
<td>WTP(q=20) (in Euro)</td>
<td>n.a.</td>
<td>24.30</td>
<td>24.87</td>
<td>24.74</td>
<td>24.60</td>
<td>25.01</td>
<td>25.06</td>
<td>25.04</td>
<td>24.40</td>
<td>26.88</td>
</tr>
<tr>
<td>WTP(q=40) (in Euro)</td>
<td>n.a.</td>
<td>37.25</td>
<td>37.88</td>
<td>37.81</td>
<td>37.56</td>
<td>38.87</td>
<td>38.99</td>
<td>38.74</td>
<td>38.08</td>
<td>43.18</td>
</tr>
<tr>
<td>WTP(q=80) (in Euro)</td>
<td>n.a.</td>
<td>39.78</td>
<td>40.40</td>
<td>40.32</td>
<td>40.48</td>
<td>41.73</td>
<td>41.91</td>
<td>41.64</td>
<td>41.26</td>
<td>46.71</td>
</tr>
<tr>
<td>WTP(q=100) (in Euro)</td>
<td>n.a.</td>
<td>41.33</td>
<td>41.89</td>
<td>41.84</td>
<td>41.28</td>
<td>43.62</td>
<td>43.78</td>
<td>43.50</td>
<td>42.96</td>
<td>49.25</td>
</tr>
<tr>
<td>Max WTP (in Euro)</td>
<td>n.a.</td>
<td>42.64</td>
<td>43.12</td>
<td>43.07</td>
<td>43.04</td>
<td>46.00</td>
<td>46.36</td>
<td>45.91</td>
<td>45.40</td>
<td>69.34</td>
</tr>
<tr>
<td>Saturation level (in hours)</td>
<td>n.a.</td>
<td>44.36</td>
<td>45.20</td>
<td>44.97</td>
<td>46.16</td>
<td>49.49</td>
<td>50.48</td>
<td>49.77</td>
<td>46.78</td>
<td>102.44</td>
</tr>
</tbody>
</table>

Notes: The 95% confidence interval appears in parenthesis (p < .05); n.a.: not applicable.
We use the estimated individual willingness-to-pay functions to calculate the average WTP for various quantities and also report the results in Table 4. Model 1, which captures only main effects, does not allow for any conclusions about the WTP for different quantities (see also Iyengar et al. 2008). In all other models, the average WTP for 20 hours of Internet access varies between 24.30 and 26.88 Euros, whereas that for 80 hours ranges between 39.78 and 46.71 Euros. The average WTP for the first 20 hours of Internet access thus is roughly 2.5 times higher than that for the second 20 hours and 7 times higher than that for the third 20 hours. These results clearly suggest that WTP substantially differs across quantity units, which highlights the finding that there is no one, single WTP. Instead, for multiple-units products or recurrently used services, WTP must be linked to the number of units consumed.

We consider all reported values for the average WTP for various quantities as face valid, with two exceptions. First, Models 5 and 9, which capture usage uncertainty on the quantity $q_{i,j}$, estimate for zero quantities a WTP of 5.71 Euro and 7.70 Euro, respectively. These WTP are higher than the WTP of any other model. The explanation for these unreasonably high values is that the direct link of usage shocks to quantity frequently leads to situations in which the quantity would become negative, if it would not be truncated at zero. This truncation increases the usage-independent WTP, captured by the parameter $c_i$. In contrast with two-part tariffs, these truncations rarely occur in three-part tariffs (Masuda and Whang 2006), because the marginal price for consumption below the usage allowance is zero, so consumption usually is well above this value. Second, the two-step estimation procedure in Model 10 provides unrealistically large values for the saturation level and the maximum WTP because it tends to overestimate the saturation level and thus the maximum WTP of those respondents who rarely choose the no-purchase option.

In contrast with Sonnier et al. (2007), we do not find major differences for the two estimation spaces, perhaps because we estimate fewer parameters for each respondent (up to 5, compared with at least 15 parameters in Sonnier et al.’s 2007 study), and we also observe more choices. Consequently, the choice of the prior distributions of WTP has only a minor impact on the posterior distributions in our study.

**Optimal Pricing**

Finally, we briefly illustrate how our models can enable a simultaneous prediction of consumers’ service purchase decisions, and their usage quantity decisions. These capabilities in turn provide a means to estimate the revenues and profits of various two-part tariffs. For ease of exposition, we focus on the optimal two-part tariff and the two special cases, that is, the optimal pay-per-use tariff (fixed fee is zero) and the flat rate (marginal price is zero). We use simulated annealing to determine the optimal prices in these tariffs, which maximize profit, and the results from Model 6 (estimation in WTP space without usage uncertainty) to model consumers’ behavior (see Masuda and Whang 2006; Schlereth et al. 2010). Thus, we neglect competition in our effort to illustrate the effects of these different tariffs better, and we assume that variable costs are .10 Euro per hour.

<table>
<thead>
<tr>
<th></th>
<th>Fixed Fee</th>
<th>Marginal Price</th>
<th>Profit</th>
<th>Revenue</th>
<th>Quantity</th>
<th>Average Quantity per Customer</th>
<th>Share of Customers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal pay-per-use tariff</td>
<td>-</td>
<td>0.86</td>
<td>-15.09%</td>
<td>-15.51%</td>
<td>-18.54%</td>
<td>-39.35%</td>
<td>+25.19%</td>
</tr>
<tr>
<td>Optimal two-part tariff</td>
<td>9.43</td>
<td>0.59</td>
<td>4,347.81 Euro</td>
<td>4,177.33 Euro</td>
<td>5,963.06 hours</td>
<td>39 hours</td>
<td>73%</td>
</tr>
<tr>
<td>Optimal flat rate</td>
<td>30.56</td>
<td>-</td>
<td>-38.46%</td>
<td>-28.39%</td>
<td>+45.09%</td>
<td>+89.35%</td>
<td>-17.16%</td>
</tr>
</tbody>
</table>

Table 5 shows that the two-part tariff naturally leads to the highest profit and highest revenue. Yet it attracts fewer customers (measured by share of customers, which is the average choice probability of all respondents) than the pay-per-use tariff. This result is intuitively appealing, because the fixed fee of the two-part tariffs will be too high for some consumers. Compared with the optimal two-part tariff, the optimal pay-per-use tariff decreases the average consumption per customer by 39.35%, which can be

---

4 We refer to Choudhary (2010) for extensions, which account for competition.
explained because of its higher marginal prices (see also Equation (13)). The flat rate induces the highest average consumption per customer but the smallest share of customers. These results again make intuitive sense, because the high fixed fee is too high for some consumers, but the inexistent marginal price encourages them to consume a lot.

In summary, the results emphasize the strong interdependency between prices and individual consumption. As we outline in Table 5, the average quantity per customer varies strongly across tariffs. As a consequence, traditional discrete choice models, such as Model 1, which cannot capture such interdependencies, are not appropriate in situations, in which consumers tend to purchase varying numbers of quantity units and value the first units more than subsequent units.

Conclusions

Two-part tariffs are prominently used by many digital services in which consumers usually purchase more than one unit, as is the case in the telecommunication, cloud computing, or banking industries. Modeling consumer choice and usage behavior with regard to such two-part tariffs is challenging because of the interdependency between prices and consumption, which requires models that allow for the simultaneous prediction of consumers' service purchase decisions, their tariff choice decisions, and their usage quantity decisions. This study is the first to propose and compare a variety of discrete choice models that capture usage behavior with two-part tariffs and that differ in their estimation space (preference versus WTP space), as well as the consideration of usage uncertainty and the complexity of parameter estimation (one-step versus two-step estimation procedure).

The results of our empirical study reveal that WTP differs significantly across quantity units and demonstrate that no single WTP exists. Therefore, the WTP per unit must be linked to the number of units consumed, which emphasize the strong interdependency between prices and individual consumption. The results also indicate that a simple (two-step) estimation procedure provides inferior results that likely overestimates saturation levels and the maximum WTP. Furthermore, different approaches for capturing usage uncertainty and different estimation spaces have only limited influence on the results. We demonstrate how our results might be used to determine optimal tariffs and illustrate that average consumption varies significantly across two-part tariffs.

Further research might extend our results in several directions. First, we only consider two-part tariffs and do not consider tariff-specific preferences, though Lambrecht and Skiera (2006) as well as Koehler et al. (2010) show that consumers tend to prefer flat rates over pay-per-use tariffs, even if they have to pay more. Strong preferences for tariffs may require estimations of tariff-specific parameters. Another limitation might be that we focused on semi-quadratic function form of the willingness-to-pay-function, even though other functional forms are also proposed in the literature, such as multiplicative, modified-exponential, or semi-logarithmic (e.g., Ater and Landsman 2013; Iyengar and Jedidi 2012). These functional forms mainly differ in whether a saturation level exist or a maximum WTP can be reached with increasing quantity. We also do not provide the respondents with additional information that would enable them to compare the stimuli more effectively, such as the quantities that would cause the different two-part tariffs to result in similar billing rates. This information might make the decisions in choice-tasks easier and lead to even better results (see Schlereth 2013 for an example).

Acknowledgments

I have to express my sincere thanks to all six reviewers and the associate editor, who provided many constructive and excellent suggestions on this project.

References


