UNIVERSITY OF CAMBRIDGE | Energy Policy Research Group

Flexible Mixed Logit with Posterior Analysis: Eliciting Willingness to Pay for Grid Resilience

EPRG Working Paper 1615

Cambridge Working Paper in Economics 1631

Laura-Lucia Richter[1] and Melvyn Weeks

Abstract This paper presents and employs an alternative approach to explore consumer preferences and willingness-to-pay (WTP) for resilience of the electricity grid. The methodological and practical relevance of this approach is demonstrated using the example of the UK's incentive regulation scheme for electricity distribution network operators (DNOs). The estimation strategy flexibly accounts for preference heterogeneity in the population, allows for scale heterogeneity (i.e. heterogeneity in the randomness of choice) and exploits individual posterior distributions to improve the estimates. Since the results suggest significant parameter heterogeneity within and across DNOs, it is argued that Ofgem's current method to evaluate consumer preferences and WTP is likely to result in an inefficient level of resilience services. The welfare implications in case of public and private goods and services are discussed. The suggested approach is straightforward to implement, could improve policy evaluations and foster more nuanced and efficient incentive regulation.

Keywords Willingness-to-Pay, electricity service quality, mixed logit, social welfare

JEL Classification C18, C38, L51, L94, Q48

Contact llr23@cam.ac.uk
Publication May 2016

[1] Contact author: Laura-Lucia Richter, Email: llr23@cam.ac.uk. Thanks for extensive comments provided by Kenneth Train and Stephane Hess.

Flexible Mixed Logit with Posterior Analysis: Exploring Willingness-to-Pay for Grid Resilience

Laura-Lucia Richter*
Faculty of Economics and
Energy Policy Research Group
University of Cambridge

Melvyn Weeks
Faculty of Economics and
Clare College
University of Cambridge

May 9, 2016

Abstract

This paper presents and employs an alternative approach to explore consumer preferences and willingness-to-pay (WTP) for resilience of the electricity grid. The methodological and practical relevance of this approach is demonstrated using the example of the UKs incentive regulation scheme for electricity distribution network operators (DNOs). The estimation strategy flexibly accounts for preference heterogeneity in the population, allows for scale heterogeneity (i.e. heterogeneity in the randomness of choice) and exploits individual posterior distributions to improve the estimates. Since the results suggest significant parameter heterogeneity within and across DNOs, it is argued that Ofgem's current method to evaluate consumer preferences and WTP is likely to result in an inefficient level of resilience services. The welfare implications in case of public and private goods and services are discussed. The suggested approach is straightforward to implement, could improve policy evaluations and foster more nuanced and efficient incentive regulation.

Keywords: Willingness-to-Pay, electricity service quality, mixed logit, social welfare

*Contact Author: Laura-Lucia Richter, Faculty of Economics, University of Cambridge, Cambridge CB3 9DD, UK. Email: llr23@cam.ac.uk. We have benefited from extensive comments provided by Kenneth Train. Chris Watts and Karl Hurley, analysts on Ofgem's Electricity Regulation team have also provided useful input.

1 Introduction

The need to replace existing infrastructure and the aim to connect an increasing share of energy from renewable sources and distributed generation to the grid will drive major investments in electricity transmission and distribution networks in the coming decades (Ofgem, 2014). The economic viability of such investments depends on a comparison of benefits and costs. As transmission and distribution networks are natural monopolies, though, there is no real market for the related services and hence no market price to signal their value. Since the 1990s regulators of infrastructure industries around the globe have therefore implemented incentive regulation models that mimic market mechanisms to promote efficiency in such natural monopolies (Jamasb et al., 2012). The challenge is to set the incentives such that the optimal level of service is provided.

To determine and incentivise the optimal service level, regulators have to estimate the marginal cost curve and a demand curve (i.e. the customers' WTP) for the respective services. Choice modelling is broadly regarded as the most suitable method for estimating consumers' WTP and thus mimic demand for service improvements (Breidert et al., 2001). In fact, the estimation of WTP via stated choice experiments has become an important part of price review processes of regulators such as Ofgem, the gas and electricity market regulator in the UK. However, usually multinomial logit (MNL) models are employed to elicit consumer preferences. These models assume homogeneous preferences within presumed consumer segments and do not allow for differences in the consumers' randomness of choice. This is despite empirical evidence of heterogeneity in preferences and WTP (Devine-Wright et al., 2010; McNair et al., 2011, see). In Ofgem's economic analyses for example customer preferences are assumed homogeneous within DNOs. In this paper we show that this is a critical and questionable assumption that is likely to lead to an inefficient level of service provision. We illustrate that if heterogeneity in valuations is not accounted for when determining the allowances, the transition of utilities to a socially efficient service level and a welfare maximising payment plan might be distorted.

We present an alternative, less restrictive approach to explore customer preferences and WTP for resilience services. We account for preference heterogeneity in the population, allow for scale heterogeneity (i.e. heterogeneity in the randomness of choice) and exploit individual posterior distributions to improve the WTP estimates and thus the economic valuations. More precisely, we start off with the estimation of a heterogeneous scale mixed logit model in WTP space, derive the individual posterior distributions based on the individuals' choices and exploit them to segment the data post estimation to calculate the total valuations per DNO. We refer to 'posterior analysis' in the sense that we analyse the conditional estimates that exploit information on the choices made by the individuals. The results of different types of posterior analysis are discussed. We demonstrate the practical relevance of this new estimation strategy based on a discrete choice experiment that informed the fifth distribution price control review (DPCR5) in the UK. In our empirical analysis the valuations for undergrounding of overhead lines and for the resilience to storms are in focus. However, the approach is equally applicable to any other stated choice experiment set up to inform incentive regulation. Our approach is straightforward to implement, could improve policy evaluations and foster more efficient incentive setting.

This paper is organised as follows. Section 2 provides background on the relevance of customer heterogeneity for incentive regulation. Section 3 presents the econometric model. Section 4 describes the data that underlies the discrete choice analysis. Section 5 provides empirical evidence of the merits of our estimation strategy. Section 6 offers some policy implications and section 7 concludes.

2 Heterogeneous Preferences for Electricity Services

The transmission system in the UK comprises over 25,000 circuit km overhead and about 2,000 circuit km underground lines. The distribution system adds further 845,000 circuit km, of which around 525,000 circuit km are underground cables and 310,000 circuit km are overhead lines (Energy Network Association, 2014). Transmission involves long distance transportation of electricity at high voltage, distribution refers to the transportation of low voltage electricity through local networks. There are two Transmission System Operators (TSOs) and fourteen Distribution Network Operators (DNOs) in the UK. They run the electricity networks and are regional monopolies. Ofgem regulates these monopoly companies and determines the allowed revenue that can be earned during the price control period. The challenge is to set the allowances such that the optimal service level is provided. TSOs and DNOs set their business plans against the price controls. More background on the incentive schemes in the UK can be found in the Appendix.

The empirical analysis of this paper focuses on incentive regulation of services delivered by DNOs. DNOs are responsible for wires and cables, for restoring the power supply in case of power cuts, connecting customers to their local network, ensuring the right voltage gets to business and consumers and for investigating any complaints that customers have regarding their electricity distribution service for example. They charge the electricity suppliers for the use of the network and the suppliers in turn charge their customers the total cost for all elements in the supply chain. In the UK the distribution component makes up about sixteen percent of a customer's bill (Ofgem, 2014). The DNOs hence have an indirect relationship with the customers and the regulator has to ensure that the DNOs don't exploit their price setting power, but charge reasonable prices. There is evidence that DNOs differ in their customer base and there is also substantial customer heterogeneity *within* DNOs. This heterogeneity exists in multiple dimensions such as location, urbanisation, grid resilience, the percentage of overhead lines, the exposure to flood and storm risks and is likely to be reflected in different valuations for resilience services. Electricity North West for example states in its investment plan for the DPCR5: 'We cover a diverse range of terrain and customer mix from isolated farms in Cumbria to areas of heavy industry, urban populations and city centers' (Electricity North West, 2008).

For several reasons the understanding of customer preferences for electricity services and how they vary across customers can inform grid infrastructure planning. Despite being essential for a reliable and resilient grid based on renewables, the related investments are often met by costly public opposition. The main issue is that more reliable and resilient networks benefit a country as a whole, but costs and benefits are not imposed on all customers to the same degree and depend on the service provided. Some investments in infrastructure impose costs such as reduced property prices, visual amenity and tourism on only few, e.g. on those households along the lines (Tobiasson et al., 2015). Other investments such as the undergrounding of overhead lines imply positive externalities for customers along the lines, but might not be equally valuable for customers further away from the lines.

In an empirical study of public beliefs about electricity supply networks Devine-Wright et al. (2010) find strong public support for placing new power lines underground, regardless of the cost.[1] Respondents currently living within sight of infrastructure were significantly more likely to strongly support undergrounding regardless of cost, in comparison to those who did not live in sight of such infrastructure. But there has been no further investigation of the WTP for undergrounding in the UK. McNair et al. (2011) were the first to explicitly estimate an average WTP for undergrounding using data from established residential areas in Canberra, Australia. They show that there is significant variation in preferences in the population, but employ a standard binary logit model due to data constraints (McNair et al., 2011). Soini et al. (2011) analyse the heterogeneity of energy landscape perceptions using a latent class method to identify different types of customers. They find significant heterogeneity of perceptions, which has implications for the planning, public communication and participatory approach related to grid infrastructure. More generally, Soini et al. (2011) find that knowledge concerning power lines significantly increased the probability of positive perceptions and reduced the probability of negative perceptions. There is also some evidence to suggest that public support for new power lines is higher when individuals believe that such lines will deliver electricity from renewable energy sources (Saint Consulting, 2009). And along the lines of previous research, the WTP for undergrounding or other kinds of investment in resilience might be higher, if customers understand that the new infrastructure serves a less carbon intense, renewable-based electricity grid. Existing research evidently supports our argument that the consideration and understanding of heterogeneity in WTP can be essential for infrastructure planning and public policy making.

Despite the evidence of heterogeneity in preferences and WTP for electricity services, in Ofgem's economic analyses customer preferences are assumed homogeneous within DNOs. We show that this is a critical and questionable assumption that is likely to lead to an inefficient level of service provision. Along the lines of Jamasb et al. (2012), who derive the welfare losses due to incorrect estimation of the marginal cost curve of quality improvements, we argue that refining demand (WTP) estimates by taking preference heterogeneity into account can increase welfare. Jamasb et al. (2012) find that the RPI-X incentive scheme did not provide sufficient incentives to improve reliability. Assuming a constant WTP they show that refined estimation of marginal cost curves for the UK electricity DNOs can increase welfare. In contrast to Jamasb et al. (2012), we take the marginal cost curve as given and illustrate how imprecise estimation of WTP can lead to welfare losses. If, for example, the regulator overestimates the value of grid resilience due to over-simplified assumptions on customer preferences, the investments in infrastructure

[1]According to estimates of National Grid undergrounding a typical 400kV double circuit power-line will cost 12 to 17 times as much as installing the same line overhead.

and thus service quality might be pushed beyond the optimal level. That is, the transition to a socially efficient service quality level and a welfare maximising payment plan might be distorted. We present a new approach to achieve more reasonable WTP estimates.

2.1 Welfare Implications of Heterogeneous Preferences

DNOs provide services with public good characteristics (such as undergrounding) as well as some services with private good characteristics (such as customer care in case of blackouts). We therefore do not only consider the efficiency and welfare implications of different assumptions on preference heterogeneity, but also investigate how the implications of our estimation strategy differ depending on the type of resilience service provided.

First, public goods and services are non-excludable and non-rivalrous in consumption, but customers might differ in their WTP for them. Undergrounding for example is a service to all customers, improving among others the resilience of the grid. But as it also improves the visual amenity, it might be particularly valuable for customers living close to the lines. Comparably, the valuation of resilience to storms might be relatively more important to customers living in areas with high storm risk.

Allocative efficiency of public goods and services provision is achieved when the social marginal benefit equals the social marginal cost i.e. $\sum_{i=1}^{N} WTP_i = MC$. The optimality condition holds regardless of the social welfare function (SWF).[2] In cases of public goods and services, correct estimation of the aggregated WTP is therefore crucial to achieve allocative efficiency: more precise estimation of the mean and hence total vertically aggregated WTP affects whether the respective public service improvement is evaluated as economically viable or not.

Distributive efficiency of public goods and services provision can be achieved by optimising the payment plan. Given the optimal quantity of the service, i.e. a pareto-optimum, there will always be a payment plan that maximises the respective SWF. The challenge is to find this plan. Different social welfare functions consider different distributions of wealth optimal. With a traditional utilitarian (Benthamite) social welfare function for example, which is simply the sum of the individual's utilities (i.e. $W = \sum_{i=1}^{N} u_i$), distributive efficiency is reached when those who value the public service most pay most for its provision. For a more general welfare function that assigns different weights (i.e. $W = \sum_{i=1}^{n} a_i u_i$) or a Rawls' welfare function for which the welfare only depends on the lowest utility (i.e. $W = min\{u_1...u_n\}$), different payment plans would be optimal. As the optimal payment plan depends on the preference distribution in the population, distributive efficiency can be improved by exploiting the individual WTP distributions - for any SWF. We hence argue that the individual WTP distributions can be exploited to inform the optimal payment/compensation plan. Since in the case of public goods and services all consumers have to consume the same quantity, the individual WTP cannot be exploited to differentiate who receives the service.

Second, regulated private services, i.e. excludable and rivalrous services, could be services such as

[2]Whether the condition is fulfilled depends on the initial distribution of wealth, if the WTP is a function of income.

customer connections to the local network, customer consultation and technical assistance in times of blackouts or the number of minutes lost in blackouts. These are all services that need not be provided to the customers to the same degree. With the development of smart grids that allow more customer-specific contracting and imply the decentralisation of the electricity system, service differentiation is likely to play a more prominent role. For such differentiable services more accurate estimation of the individual WTP distributions can improve both allocative and distributive efficiency: the aggregate demand curve now results from a *horizontal* aggregation of the individual valuations. While the homogeneous valuations assumed in the MNL model imply a perfectly elastic demand curve, the consideration of heterogeneous individual WTPs allows for a downward sloping demand curve, which is illustrated in figure 1. The WTP is a downward sloping function of service quality.

Allocative efficiency of private goods and services is achieved when the horizontally aggregated WTP equals the MC. Correct estimation of the individual level WTP is therefore important to achieve allocative efficiency: more precise estimation of the individual level and hence horizontally aggregated WTP affects whether the respective service level is economically viable or not.

Distributive efficiency is achieved when the SWF is maximised. An analysis of the individual WTP distributions makes it possible to not only determine the efficient level of output, but also to provide them such that the SWF is maximised (i.e. inform who should receive the service and who should pay for it). Providing the same service quality to all customers is not necessarily welfare maximising. If agents have heterogeneous preferences, they will want to trade away from a uniform distribution, which is an argument for a differentiation of service quality. While some consumers might ceteris paribus accept a lower degree of supply reliability (e.g. because they hardly use any electric appliances at home), others might be willing to pay for a higher degree of supply reliabilty. Another example is the service provided by call centers: while some customers might value the availability of customer support, others might not be willing to pay for it. Thus, in case of private services, the individual WTPs can be exploited to determine not only who should pay for the services, but also who should receive them.

In this context it needs to be considered that the WTP for service quality is likely to be positively correlated with income. Providing the goods and services preferrably to those with the highest WTP is therefore prone to social equality concerns. To address such potential social equality concerns, a minimum level of service, e.g. for fuel-poor households, should be guaranteed. The choice of the social welfare function shall not be the focus of this paper. However, conditional on income, services with private good characteristics could be provided to those who are willing to pay most, which could improve distributive efficiency.

In summary these considerations illustrate that not only the social optimal level of service might be distorted, but that also the payment plan and the distribution of the services (in the case of regualated excludable and rivalrous services) is likely to be inefficient, if customer heterogeneity is not taken into account. In our empirical analysis we thus distinguish between implications of preference heterogeneity for allocative and distributive efficiency in case of both regulated public and private services. In the case of public services only the aggregate WTP matters for allocative efficiency and the individual level WTPs

Figure 1: Welfare Loss from Misspecification of Demand (WTP)

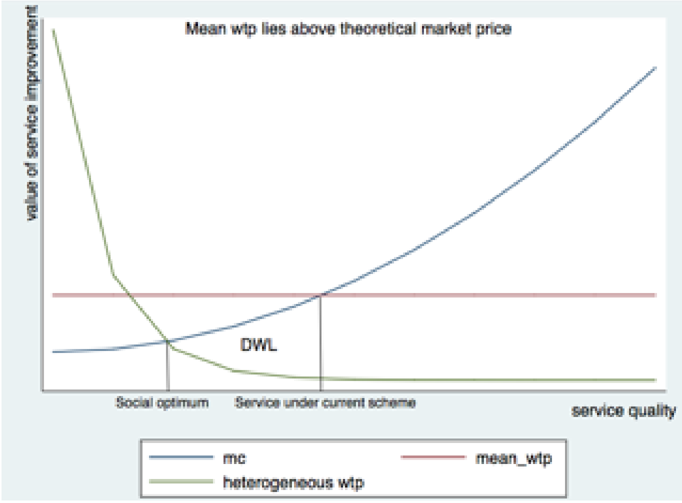


Figure 1 illustrates the potential welfare loss from misestimation of WTP in case of regulated private goods. The horizontal summation of heterogeneous individual WTPs leads to a downward sloping, rather than totally elastic, demand curve. As the efficient service level lies where demand intersects with marginal cost, misspecification can lead to inefficient service levels. Moreover, distributional efficiency can be distorted, if the heterogeneous valuations are ignored.

can mainly inform the welfare maximising payment plan. In the case of private services on the other hand, the individual WTPs can improve both allocative and distributive efficiency.

3 Econometric Model

In the following three subsections we introduce the econometric model underlying our estimation strategy. We first present a mixed logit model that is flexible in the sense that it can accommodate random preference and scale parameters and allows for different types of scaling. We then introduce the concept of WTP space models. In the last part of this section we discuss the potential of posterior analysis to refine and explore WTP estimates and their heterogeneity in the context of different types of electricity services.

3.1 Flexible Mixed Logit Framework

In discrete choice modelling the researcher estimates the probability (rather than quantity) of choice using the random utility model (RUM). The aim is to study how multiple consumer and product attributes jointly affect choices and to estimate implicit prices not only for the bundled service, but also for its different attributes. Consider N consumers $i = 1, 2, 3, ..., N$, J choice alternatives $j = 0, 1, 2, 3, ..., J$ and T choice situations $t = 1, 2, 3, ..., T$. In the RUM consumer i chooses product j to maximize his utility that is affected by L individual and K product characteristics. Let there be unobserved heterogeneity that can capture why customers with similar observable characteristics might still make different choices and why products with similar observable characteristics might still have different market shares. Let d_{ij}

be an indicator for consumer i choosing alternative j. The probability that customer i buys alternative 1 then

$$\Pr(d_{i1} = 1) = \Pr(U_{i1} > U_{i0} \cap ... \cap U_{i1} > U_{iJ}) \tag{1}$$

Let the utility be separable in price and non-price attributes and consider the following traditional utility specification underlying the logit model in preference space:

$$U_{ijt} = \alpha p_{jt} + \boldsymbol{\omega}' \mathbf{v}_{jt} + \epsilon_{ijt} \tag{2}$$

$$\alpha = \alpha_0 + \boldsymbol{\alpha}_1 \mathbf{x}_{it} \tag{3}$$

$$\omega_k = \omega_{0k} + \boldsymbol{\omega}'_{1k} \mathbf{x}_{it} \quad \forall k = 1...K \tag{4}$$

where p_{jt} measures the price or cost of alternative j and $\mathbf{v}_{jt}$ is a $(K \times 1)$ vector of observable non-price attributes. α and $\boldsymbol{\omega}$ are vectors of attribute coefficients to estimate. ϵ_{ijt} is an independent and identically distributed (*iid*) random term that captures unobserved characteristics and follows a standardised type I extreme value distribution with constant variance $Var(\epsilon_{ijt}) = (\pi^2/6)$. While in a traditional multinomial logit (MNL) framework the individual's choice only depends on individual characteristics, $\mathbf{x}_{it}$, the conditional logit (CL) framework also includes alternative specific characteristics, $\mathbf{v}_{jt}$, in the utility function. However, many papers, including ours, refer to the CL model as the MNL model with the understanding that individual and alternative specific characteristics are allowed in the response probability (Wooldridge, 2010).

A more generalised utility specification in preference space is given as

$$U_{ijt} = \alpha_i p_{jt} + \boldsymbol{\omega}'_i \mathbf{v}_{jt} + \sigma_i^{-1} \epsilon_{ijt} \tag{5}$$

$$\alpha_i = \alpha_0 + \boldsymbol{\alpha}_1 \mathbf{x}_{it} + \nu_{1it} \tag{6}$$

$$\omega_{ik} = \omega_{0k} + \boldsymbol{\omega}'_{1k} \mathbf{x}_{it} + \nu_{2ikt} \quad \forall k = 1...K \tag{7}$$

where α_i and $\boldsymbol{\omega}_i$ are now individual specific vectors of attribute coefficients to estimate. The error term is scaled and has variance $Var(\sigma_i^{-1}\epsilon_{ijt}) = (\pi^2/6\sigma_i^2)$. $1/\sigma_i$ can capture individual heterogeneity in the error variance. $(\nu_{1it}, \boldsymbol{\nu}_{2it})$ are random components that follow a multivariate distribution to be specified by the researcher and capture unobserved individual characteristics. Equation 5 nests several models that differ with respect to their flexibility to accommodate parameter heterogeneity.

Imposing the restrictions $\nu_{1i} = \boldsymbol{\nu}_{2i} = 0$ and $\sigma_i = 1$ on equation 5 leads back to equation 2, which is the utility specification underlying the MNL model: the vectors α and $\boldsymbol{\omega}$ are restricted to be homogeneous conditional on $\mathbf{x}_{it}$ and the error variance is restricted to be constant across individuals and equal to $Var(\epsilon_{ijt}) = (\pi^2/6)$. As shown above, the MNL model can introduce heterogeneity in the marginal effects of observed product attributes, so-called preference heterogeneity, by including interactions of attribute and individual characteristics (e.g. $\mathrm{v}_{kjt} \times \mathbf{x}_{it}$) into the utility function. One major challenge is, though, to determine which demographics matter for which attributes and in which combination. While the log-likelihood of different specifications can provide insights which demographics are relevant, it is often infeasible to consider all possible interactions. And even if all relevant interaction terms were included, this would imply a loss of a large number of degrees of freedom. A second way to introduce preference heterogeneity into the MNL framework is to presume specific customer groups and estimate segmented MNL models on those. Drawbacks of this approach are subjectivity with respect to the segment definition and small sample sizes that hamper robust estimation.

A less restrictive model than the MNL model is a random parameter model that accommodates preference heterogeneity resulting from both observable and unobservable alternative characteristics, but still constrains the error variance to be constant, i.e. $\sigma = 1$. In such a mixed logit (MXL) model the preference parameters α_i and $\boldsymbol{\omega}_i$ are functions of observable individual specific characteristics (as before) and of an additional unobserved random error term, ν_{ikt} that follows a continuous distribution. These mixing distributions are to be specified by the researcher. While the MXL model can accommodate any kind of multivariate distribution of the preference parameters, analysts frequently impose univariate normal distributions. The price parameter is often assumed to follow a log-normal distribution. The researcher then estimates the hyper-parameters of the coefficient distributions such as mean and variance.[3] In such MXL models, customers with the same observable characteristics (i.e. conditional on $\mathbf{x}_{it}$) may hence still value product characteristics differently.

Especially when estimating WTPs, a normality assumption imposed on the marginal utility coefficients in preference space comes along with problems resulting from the unbounded nature and the symmetry of the normal distribution. Hess (2010) points out that it is often not clear whether the findings actually reflect the true underlying heterogeneity in the data or are simply a result from the distributional assumptions. More flexible mixing distributions that do not rely on strict symmetry assumptions and allow bounds on either side can have significant advantages over more assumption-bound counterparts. More flexible models are often more difficult to estimate, though, with convergence issues and problems with parameter significance (Hess, 2010). When choosing the appropriate model researchers thus have to trade-off the cost and benefits of working with more flexible mixing distributions. However, all parametric models rely on distributional assumptions which can lead to considerable estimation bias, if they are not fulfilled Fosgerau (2006).

An alternative to imposing parametric distributions is the use of semi- or non-parametric distributions. While usually more computationally costly, such models are less restrictive and have recently gained increasing attention from choice modellers (e.g. Garcia (2005); Fosgerau (2007); Li (2011); Fosgerau (2014)). Fosgerau (2007), for example, uses a range of non-parametric techniques to examine the specification of a model to evaluate the WTP for travel time changes. Fosgerau (2014) refers to non-parametric approaches in the sense that the discription of some unknown distribution is non-parametric and is embedded in an otherwise parametric model, with this combination hence being semi-parametric.

[3] In the case of a multivariate normal distribution, $(\nu_{1i}, \boldsymbol{\nu}_{2i}) \sim MVN(0, \Sigma)$ where Σ denotes the covariance matrix of the preference parameters α_i and $\boldsymbol{\omega}_i$.

He highlights that there are situations in which it is not desirable to impose a specific functional form, e.g. if the distribution of WTPs is to be inferred from the data. And Li (2011) points out that although the Gumbel distribution is a good approximation in some applications, it is chosen mainly for mathematical convenience and can be restrictive in many scenarios. He relaxes the assumption of the Gumbel distribution to include a large class of distributions that allow heteroscedastic variances and derives a semi-parametric choice model. We do not perform any non-parametric estimation, but this could be part of further research.

Instead, we test parametric mixed logit models that accommodate mixing distributions that differ in their flexibility and parameter space. As mentioned above, the mixed logit model can accommodate mixing distributions of any kind. While researchers usually assume homogeneous error variances, the model can also accomodate a utility specification that allows the error variance to vary across individuals i, with $Var(\sigma_i^{-1}\epsilon_{ijt}) = \pi^2/6\sigma_i^2$. The heterogeneous error variance is motivated by the argument that a significant part of the heterogeneity retrieved in random coefficient models might actually be due to scale rather than preference heterogeneity. Sources of scale heterogeneity can be manifold, among them heterogeneous experience with the object of interest that leads to heterogeneous randomness of choice. In our application, previous experience with power cuts due to extreme events such as floods or storms is likely to reduce the randomness of customers' choices for example. The unobserved σ_i can capture such factors that are uncertain from the decision maker's view rather than only from the researcher's perspective. A random scale parameter is thus conceptually different from a random preference parameter.

A simple reparameterisation, multiplying equation 5 by σ_i, yields a model that allows for separate random preference and scale parameters:

$$U_{ijt} = \underbrace{(\sigma_i\alpha_i)}_{\lambda_i}p_{jt} + \underbrace{(\sigma_i\boldsymbol{\omega}_i')}_{\mathbf{c}_i'}\mathbf{v}_{jt} + \epsilon_{ijt} \tag{8}$$

As in equation 2 the idiosyncratic error term in this parameterisation follows an extreme value type I distribution with constant and standardised variance again ($Var(\epsilon_{ijt}) = \pi^2/6$), the model in equation 8 can hence be estimated like the MXL model in equation 5. In specification 8 the attribute coefficients are now the product of two independently distributed separate coefficients. However, while the existence of scale heterogeneity has been widely acknowledged in the literature, specification 8 also illustrates that α_i and $\boldsymbol{\omega}_i$ cannot be separately identified from σ_i (Hess and Rose, 2012). Some normalisation is therefore necessary.

To illustrate the identification problem we follow Hess and Rose (2012). In a case in which λ_i and $\mathbf{c}_i$ are considered attribute coefficients, they are directly estimated based on a mixed logit framework where scale is normalised to 1 ($\lambda_i = \alpha_i$ and $\mathbf{c}_i = \omega_i$). If there *was* scale heterogeneity, this model that just considers λ_i and $\mathbf{c}_i$ could not distinguish the variation resulting from scale from preference heterogeneity; they are confounded. Even a model that considers the attribute coefficients as products of two components that follow separate distributions ($\sigma_i \times \alpha_i$) and ($\sigma_i \times \boldsymbol{\omega}_i$) does not separately identify the preference and scale parameters. In fact, a model that allows for correlation between λ_i and c_i is structurally equivalent to a model that considers α_i, $\boldsymbol{\omega}_i$ and σ_i separately, provided that λ_i and $\mathbf{c}_i$ follow the same distributions as the products $(\sigma_i \times \alpha_i)$ and $(\sigma_i \times \boldsymbol{\omega}_i)$, respectively.

The so-called generalised multinomial logit (GMNL) model, first proposed by Keane and Wasi (2013) and operationalised by Fiebig et al. (2010) and Hensher and Greene (2011), incorporates a separate random scale parameter in addition to the random preference parameters. Distributions are imposed on both preference and scale coefficients. Fiebig et al. (2010) extend the parameterisation of the individual specific parameters in the GMNL model from $\mathbf{c}_i = (\sigma_i \boldsymbol{\omega}_i)$ to

$$\mathbf{c}_i = \sigma_i \bar{\boldsymbol{\omega}} + [\gamma + \sigma_i(1-\gamma)]\boldsymbol{\nu}_i \tag{9}$$

$$\sigma_i = \exp(\bar{\sigma} + \tau\epsilon_{0,i}) \tag{10}$$

where $\boldsymbol{\nu}_i$ follow a multivariate normal distribution ($\boldsymbol{\nu}_i \sim MVN(0, \Sigma)$). $\epsilon_{0,i}$ follows iid standard normal distributions such that the parameter σ_i is log-normally distributed ($\sigma_i \sim logN(\bar{\sigma_i}, \tau)$) where $\bar{\sigma}$ and τ are the location and scale parameter respectively. The mean of this log-normal distribution is $E(\sigma_i) = \exp(\bar{\sigma} + \frac{\tau}{2})$. Therefore setting $\bar{\sigma} = -\frac{\tau^2}{2}$ normalises $E(\sigma_i)$ to 1, which is done for identification. A parameter τ significantly different from zero indicates significant heterogeneity in σ_i. γ is a weighting parameter that determines in how far only the mean ($\bar{\boldsymbol{\omega}}$) or also the random component of the preference parameter (ν_i) is scaled. The full GMNL model, where $\gamma \in [0, 1]$, allows for differential scaling of β and $\boldsymbol{\nu}_i$. If $\gamma = 1$ and $\tau \neq 0$, only the mean preference parameter is scaled (i.e. $\mathbf{c}_i = \sigma_i \bar{\boldsymbol{\omega}} + \boldsymbol{\nu}_i$). This model is called GMNL-I and adopts the prior that the individual level parameters $\mathbf{c}_i$ follow a mixture of normals with different means but equal variances (namely the variance of $\boldsymbol{\nu}_i$). In GMNL-II, $\gamma = 0$ and $\tau \neq 0$ and hence $\mathbf{c}_i = \sigma_i(\bar{\boldsymbol{\omega}} + \boldsymbol{\nu}_i)$ so that the standard deviations are proportional to scale. GMNL-II adopts the prior they are a mixture of normals with proportionally different means and standard deviations (Fiebig et al., 2010). This view is consistent with the traditional view of scale that considers the entire utility function being scaled. Whenever $\gamma > 0$, the scale parameter does not represent scale in its traditional sense, which is an additional argument to not claim identification of scale heterogeneity. Instead, we emphasise that GMNL is a specific form of the mixed logit specification with flexible mixing distributions. And GMNL still accommodates the possibility to impose the necessary restrictions to achieve low computational costs and convergence, in our case $\gamma = 1$.

As Hess and Rose (2012) show, various model specifications can merely be seen as different parameterisations of the mixed logit model and any gains in fit obtained in random scale models are the result of more flexible distributions than they are frequently imposed. As an example, the GMNL specification nests the logit models discussed above. If $Var(\boldsymbol{\nu}_i) = 0$ and $\sigma_i = 1$, the MNL model results. If $Var(\boldsymbol{\nu}_i) = 0$ and $\sigma_i \neq 1$ the model turns into the so-called scaled multinomial logit (S-MNL) model. This leads to the case mentioned above, where the preference parameters α and $\boldsymbol{\omega}$ are assumed constant and uncorrelated, but λ_i and $\mathbf{c}_i$ are by construction perfectly correlated through σ_i. Any parameter heterogeneity discovered might result from preference heterogeneity that the model fails to capture. Vice versa, if λ_i and $\mathbf{c}_i$ are independent, the scale parameter is assumed to be constant, leading back to a MXL model without heterogeneous scale. In such a model any parameter heterogeneity discovered might result from scale heterogeneity that the model fails to capture. If λ_i and the components of $\mathbf{c}_i$ vary with less than unit correlation, α_i and $\boldsymbol{\omega}_i$ are necessarily varying in addition to scale, but the types of variation cannot be disentangled. We follow the interpretation of Rose et al. (2012) that the GMNL formulation allows the attribute coefficients, here λ_i and $\mathbf{c}_i$, to follow flexible mixtures of distributions (e.g. of log-normal and normal distribution), which does not necessarily imply separate identification of preference and scale. However, as we argue below, if the model is correctly specified, the estimation in WTP space rather than in preference space can yield unconfounded WTP estimates, while the price coefficient remains confounded by scale.

3.2 Models in WTP Space

If the mixed logit model is specified in preference space, the distributional assumptions are imposed on the preference (and scale) parameters. The WTP for an attribute j is derived from the ratio of that attribute's parameter to the price parameter. The ratios of the random parameters can produce extreme and unreasonable ranges of valuations and might bias the estimated means (Hess, 2007; Daly et al., 2012). Depending on the distribution imposed on the coefficients, the WTP distributions may or may not have finite moments. Daly et al. (2012) present a theorem that allows researchers to test whether the moments of the WTP distribution exist and suggest several ways to ensure finite moments of WTPs.

One straight forward way to address the issue of undefined moments of WTP and ensure finite moments is to reparameterise the model such that the distributional assumptions can be imposed directly on the WTPs and their moments estimated directly from the data as suggested by Train and Weeks (2004). Rearranging equation 8 and re-parameterising yields the heterogeneous scale mixed logit model in WTP space:

$$U_{ijt} = (\underbrace{\sigma_i \alpha_i}_{\lambda_i})[p_{jt} + (\underbrace{\boldsymbol{\omega}_i'/\alpha_i}_{\mathbf{w}_i'})\mathbf{v}_{jt}] + \epsilon_{ijt} \tag{11}$$

This model is theoretically equivalent to the model in preference space, but the re-parameterisation implies differences in the ease of implementation of different distributions. In WTP space, for example, it is straightforward to impose normal distributions on the WTPs. Any differences in model fit compared to models estimated on the same data in preference space are mainly a result of the distributional assumptions imposed on the parameters. If $(\lambda_i \mathbf{w}_i)$ in WTP space follows the same distribution as $(\sigma_i \boldsymbol{\omega}_i)$ in preference space, model fit should be the same.

Equation 11 also illustrates that the scale parameter σ_i does not directly impact the WTPs, but is picked up separately by λ_i, i.e. by the price coefficient in WTP space. Differences in WTP estimates between models that do and do not include a random scale parameter are thus the consequence from any impacts that the inclusion of the scale parameter has on the remaining model parameters (Hess, 2007), in our case on the price coefficient in WTP space. λ_i incorporates any differences in scale across respondents

(Train and Weeks, 2004). α_i and σ_i are not separately identified. Any changes in the distribution of $\lambda_i = \alpha_i \sigma_i$ implicitly impacts the distribution of the preference parameters which in turn impacts model fit.

As in preference space, specification 11 nests multiple models that depend on the restrictions imposed on the parameters. We impose the constraints on the coefficients as discussed and listed in table 1. The first column of table 1 lists the model names, the second column the equivalent names of the nested models, column three indicates whether the model constraints the scale parameter to be constant ($\sigma = 1$) or not (σ_i). The fourth column indicates whether the model allows for random preference parameters ($Var(v_i) = 0$) or not ($Var(v_i) \neq 0$) and the last column summarizes the model by listing the expected utility in WTP space.

Model M2, the MNL model in WTP space, can be obtained by imposing the restrictions $\sigma = 1$ and $Var(\nu_i) = 0$. Since the scale parameter is normalised to 1, α can be retrieved from the coefficient of the constant. In M3, which is equivalent to the S-MNL model, the constant picks up $\lambda_i = (\sigma_i \alpha)$. With the normalisation of Fiebig et al. (2010) that $E(\sigma_i) = 1$, α can be derived in expectation: $E[\sigma_i \alpha] = \alpha$. M4, the MXL model in WTP space, on the contrary allows for random preference parameters, but imposes the restriction $\sigma_i = 1$. As in MNL, α_i can be backed out from the constant. The specification in 11 also nests models that allow for correlated preference coefficients. Since in such models $Cov(\nu_{2ikt}\nu_{2ilt}) \neq 0$ for any k,l the correlation of preference parameters translates into a correlation of WTP estimates. If all covariances are unconstrained, this requires estimation of a large number of additional parameters and comes along with the loss of a high number of degrees of freedom.

Assuming that the *true* model does include a separate random scale parameter as in equation 11, i.e. with $E(U) = (\sigma_i \alpha_i)[p_{jt} + (\boldsymbol{\omega}_i'/\alpha_i)\mathbf{v}_{jt}]$ and the idiosyncratic error follows an extreme value type I distribution, the estimates resulting from the models that constrain either of the parameters to be constant, will be confounded due to misspecification in at least one dimension. When assuming a constant preference parameter α for example, any variation captured by σ_i can also result from variation in α_i that the constrained model fails to capture. If the model is misspecified, the distributional assumptions imposed on the parameters are wrong. λ_i and c_i will in truth follow the product of the two distributions of σ_i and α_i and σ_i and $\boldsymbol{\omega}_i$ respectively. Moreover, if a relevant scale parameter is omitted, a homoscedastic variance is imposed on a heteroscedastic one. This results in confounded mean estimates, as the extreme value type I distribution is necessary for unconfounded estimation.

Table 1: Models in WTP Space - Summary

Model	Equivalent	σ_i	$Var(\nu_i)$	$E(U_{it})$ in GMNL form
M2	MNL	1	0	$(\sigma\alpha)[p_{jt} + (\boldsymbol{\omega}'/\alpha)\mathbf{v}_{jt}]$
M3	S-MNL	σ_i	0	$(\sigma_i\alpha)[p_{jt} + (\boldsymbol{\omega}'/\alpha)\mathbf{v}_{jt}]$
M4	MXL	1	$\neq 0$	$(\sigma\alpha_i)[p_{jt} + (\boldsymbol{\omega}_i'/\alpha_i)\mathbf{v}_{jt}]$
M5	GMNL-I	σ_i	$\neq 0$	$(\sigma_i\alpha_i)[p_{jt} + (\boldsymbol{\omega}_i'/\alpha_i)\mathbf{v}_{jt}]$

The table lists all models in WTP space. Model M2 is the baseline model in WTP space, M3 only allows for heterogeneous scale, M4 only for preference heterogeneity, M5 allows for preference and scale heterogeneity. We constrain $\gamma = 1$ in all models, reflecting a prior that the individual level parameters follow a mixture of normals with different means but equal variances.

3.3 Conditional Distributions

Our approach is based on the premise that there exists heterogeneity in preferences and valuations for grid resilience across customers. This provides direct motivation for the use of a random coefficient model that combines a standard conditional logit kernel with a mixing distribution of any kind (e.g. log-normal, normal or the product of both). In the first step of the estimation we abstract from covariates and estimate the hyperparameters of the preference distribution in the population for each attribute. Letting ω denote a vector of random parameters, let $g(\omega|\eta)$ represent this distribution, with hyperparameters η.

A limitation of such an approach that relies on the unconditional distributions is that they represent random taste heterogeneity in the population. This implies that although relative to the conditional logit model we have information on the variance around mean parameters, we are not able to locate any individuals within the distribution. Following Train (2003) and further work by Hess (2010) and W.H. Greene (2005), analysts have addressed the related limitations by working with the conditional distributions, say $g(\omega|\mathbf{y}_i)$, where $\mathbf{y}_i$ is a $T \times 1$ vector denoting the sequence of choices made by individual i. Conditional distributions allow to infer the most likely position of each sampled individual on the distribution of sensitivities or valuations exploiting the information on their choices made. Let the conditional density $g(\omega|\mathbf{y}_i)$ be

$$g(\omega|\mathbf{y}_i) = \frac{L(\mathbf{y}_i|\omega)f(\omega|\eta)}{\int_\omega L(\mathbf{y}_i|\omega)f(\omega|\eta)\,\mathrm{d}\omega}, \tag{12}$$

where $L(\mathbf{y}_i|\omega)$ denotes the likelihood of observing this sequence of choices given ω. Since the denominator of 12 has no closed form solution, estimation via maximum likelihood is not possible. However, the conditional distributions can be simulated. Replacing (12) by a discrete approximation, the conditional mean for consumer i for example may be written as

$$\widehat{E_i(\omega)} = \frac{\Sigma_{r=1}^R L(\mathbf{y}_i|\omega_r)\omega_r}{\Sigma_{r=1}^R L(\mathbf{y}_i|\omega_r)}, \tag{13}$$

where ω_r are independent multi-dimensional draws from $f(\omega|\eta)$ at the estimated values for η. This conditional mean can be interpreted as the most likely value for a consumer i whose choices y_i were observed. Our estimation strategy utilises the conditional distribution $g(\omega|\mathbf{y}_i)$ to compute customer-specific estimates of mean WTPs and then aggregates these by DNO. We thus derive the conditional means of the coefficient distribution for the sub-group of individuals who face the same alternatives and make the same choices (Train, 2003).

There are a number of advantages of conditional distributions both in preference and WTP space. Firstly, since the number of choices made is finite, outlier problems can be reduced (Hess, 2010). This is not only important when estimating WTPs in preference space where the WTPs are derived as ratios from estimated coefficients, but also in WTP space, where potentially unbounded mixing distributions can result in misleading estimates. Further merits of using the conditional distributions are that posterior analysis can be conducted, linking the conditional estimates back to socio-demographics and that more substantial tests of coefficient correlations are possible (Hess, 2010). We perform classical simulation rather than Bayesian, but refer to 'posterior analysis' in the sense that we explore the conditional estimates derived based on the individuals' choices. We relate the conditional mean estimates back to individual characteristics with the aim to shed light on different consumer types in the sample. In this context we distinguish two components of the total variance of the conditional distribution: the variance of the individual-level conditional means (the between variation) and the variance around these means (i.e. the within variance). If the between variance captures a sufficiently large share of the total variation in a coefficient, the individual conditional means have potential to be useful in distinguishing customers (Train, 2003). Train (2003) discusses the possibility to employ cluster analysis based on conditional mean estimates in order to respecify the model towards different segmentation.

We combine the advantages of the estimation in WTP space with the merits of the conditional distribution $g(\omega|\mathbf{y}_i)$ to compute the aggregate WTPs by DNO. GMNL in WTP space as specified in (11) serves as a flexible main model framework as it nests models with and without preference and scale heterogeneity, can accommodate correlated coefficients and allows to impose distributional assumptions directly on the WTPs. Moreover, it allows for the derivation of individual-specific posterior distributions, the individual posterior means in particular. This can shed light on the different segments and sources of heterogeneity without assuming them a priori. In our empirical application the key aim of the individual level posterior analysis is to refine the estimated aggregated valuations per DNO and to exploit the distributions for more effective payment plans and distribution of services.

An alternative approach to the use of conditional distributions proceeds by extending the conditional logit model by utilising a DNO indicator to construct a linear index which allows for DNO specific mean parameters. This approach accommodates variation in WTP between DNOs but not within. We note that in a world in which the regulator faces a large sample for *each* DNO (e.g. 1,000 observations or more) it would be possible to combine this multiplicative index with a random coefficient model. Although such an approach has a number of advantages, the cost of the respective stated choice experiments would be significantly higher and we note that such a model generates a proliferation of both mean and covariance parameters.

4 Data

Discrete choice modelling is useful for the estimation of the WTP for goods and services, to derive demand and consumer surplus as well as for optimum pricing. In fact, the exploration of customer preferences and estimation of WTP via stated choice experiments has become an important part of price review processes of regulators such as Ofgem in the UK. To demonstrate our suggested new approach to estimate reasonable WTP distributions we use the stated preference survey 'Expectations of DNOs & Willingness to Pay for Improvements in Service' (Accent, 2008), which was conducted for Ofgem in 2008 and informed the price control period from 2010 until 2015. One of the objectives was to determine domestic and business customer priorities and WTP for investments by the DNOs. The survey was conducted with 2,002 customers belonging to the fourteen British DNOs (around 150 customers per DNO). Customers were explained what DNOs are and what they do. They were shown a showcard illustrating that DNOs own the wires and cables, have the duty to connect any customer requiring electricity within their area, maintain the connection and must maintain an efficient cost-effective and coordinated system to distribute electricity, e.g. the overhead lines. They were then confronted with three choice experiments that each included six choice tasks related to grid reliability and resilience. A questionnaire accompanying the experiments asked for (rather few) further customer characteristics: DNO, rural or urban living environment, age, income, electricity usage level, bill, power cut experience.

Table 2: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
DNO	-	-	1	14	2,002
Annual Electricity Bill (£)	527.58	382.81	24	8400	2,002
Dummy Rural Area	0.266	0.442	0	1	2,002
Dummy Fuel Poor	0.157	0.364	0	1	2,002
Dummy Power Cut Experience, past 3 yrs	0.516	0.500	0	1	2,002
Dummy Unplanned Cuts (past 12 months)	0.462	0.499	0	1	2,002
Number Unplanned Cuts (past 12 months)	0.897	1.636	0	30	1,985
Length (min) Unplanned Cuts (past 12 months)	88.277	264.234	0	4320	2,002
Dummy Planned Cuts (past 12 months)	0.073	0.260	0	1	2,002
Number Planned Cuts (past 12 months)	0.087	0.353	0	5	1,999
Length (min) Planned Cuts (past 12 months)	6.347	39.007	0	480	1,988

Table 2 (and H.1 in the Appendix) summarise the main statistics of the 2,002 customers in the sample.[4] The average annual electricity bill across customers and DNOs was £527, but varied from a £24 to £8,400. Around a quarter of customers indicated to live in rural areas and about 16 percent were fuel poor (i.e. their energy expenditures made up more than ten percent of their income). The median social class was class C1 and 81 percent of the customers were class C1 or below. The median income lay between £20,000 and £30,000 (see table H.1 in the Appendix). 30 percent had children under five years and 60 percent of the sample were under 50 years old. 51 percent of the sample had some power cut experience in the past twelve years (around 46 percent experienced unplanned cuts and only seven percent experienced planned cuts). The number of experienced unplanned cuts in the past 12 months ranged from zero to 30, with an average of 0.90. The number of planned cuts only ranged up to five and had a low average of 0.09. The length of the unplanned cuts was on average 88 minutes, but reached

[4]We exclude EDF London from our analysis, as the experiment differed for this DNO.

up to 4320 minutes (i.e. 72 hours). The median length of experienced unplanned blackouts was zero, the 75 percentile was 45 minutes and 10 percent of the sampled customers experienced unplanned cuts that lasted longer than 3 hours. The average length of planned cuts was 6 minutes and the maximum experienced length was 480 minutes (i.e. 8 hours). These numbers indicate that most customers in the representative sample had no or little experience with blackouts, but that there were some customers who experienced remarkable interruptions of services.

Since Accent's (2008) study consisted of three separate experiments on resilience services, packaging effects must be considered: the three lower level experiments can lead to an excessive value of WTP for all improvements (due to so-called budgeting or halo effects). Accent (2008) found that Experiment 3 needed smaller package adjustment than the other experiments, which makes it best suited for separate analysis. We thus use Experiment 3 to explore customer preferences and WTP for resilience of the electricity grid and to illustrate the improvements that our approach can deliver compared to the workhorse models used by regulators to date.

The attribute levels in Experiment 3 were defined as deviations from the status quo and are listed in table 3. The four considered non-price service attributes are: (1) commitment to undergrounding of overhead lines in areas of outstanding natural beauty and national parks for amenity reasons (with a status quo of 1.5 percent per annum across the UK and levels ranging from zero to five percent per annum), (2) the number of customers affected by blackouts due to major storms (with DNO specific status quo and levels ranging from twenty percent less to twenty percent more customers being affected), (3) the number of major electricity sites across the UK exposed to potential flood risk (with status quo of 1,000 sites and levels ranging from 1,000 to 850 sites)[5] and (4) the investment (by distributor) to current mobile generation equipment and vehicles to reduce CO2 emissions (with levels indicating zero, five or ten percent of replacement). The change in cost was also defined relative to the current customer bill and the absolute change (reflecting the change in the distribution component) ranged from a bill reduction by £183 to an increase by £457. An example of a choice card can be found in figure 5 in the Appendix.

Table 3: Attributes and levels as defined for the choice experiment 3 (Accent, 2008).

Attribute	# of levels	levels
Commitment to undergrounding	4	0%, 1.5% (as now), 3%, 5%
Customers affected by major storms	5	+20%, +10%, as now, -10%, -20%
Sites exposed to flood risk	4	1,000 (as now), 950, 900, 850
Investment to reduce CO2 emissions	3	as now, replace 5%, replace 10%
% change in annual electricity bill[6]	9	+30, +25, +20, +15, +10, +5, 0, -5, -10

Table 3 lists the attributes and levels of resilience services as considered in Experiment 3 (Accent, 2008). All changes in the attribute levels were defined relative to the status quo.

As described above, many of the attribute levels were based around the status quo service level for a particular DNO. Since the general public was considered mostly unfamiliar with their electricity distribution service quality, one of the choice alternatives showed all attributes at their current levels and the respondents current bill. In this way, respondents were reminded of what service they currently

[5] An increasing environmental risk with climate change is flooding from river or coastal sources. Since many major sites are located close to the water for cooling purposes, programs to improve resilience to floods are increasingly needed.

pay for. This inclusion of the status quo alternative was also meant to bring more realism to the choice exercises, allowing respondents to indicate that they would stay with their current service (Accent, 2008). Nevertheless, while not the focus of this research, several potential shortcomings could result from this experiment. As it is a *stated* choice experiment on electricity services that have relatively abstract public good characteristics, strategic and/or hypothetical biases might impact the WTP estimates. Some consumers might strategically make their choices to indicate a low WTP for resilience in order to avoid potential future tax increases for example, leading to strategic bias. And hypothetical bias might result from the abstract nature of the experiment, its attributes and levels. Moreover, the experimental design could be improved by implementing d-efficient rather than fractional factorial design for example. Efficient designs aim to result in data that generates parameter estimates with as small as possible standard errors. If some prior information about the parameters is available (e.g. from a pilot study), the asymptotic variance-covariance matrix can be determined. Hence, whenever there is any prior parameter information available (even if just the signs of the parameter), the experimental design could be improved by the use of an efficient design.

5 Empirical Analysis

5.1 Model Specification

We start off with the premise that the customer base is not only substantially heterogeneous across but also within DNOs. This heterogeneity can exist in multiple dimensions such as location, urbanisation, resilience of the grid, the percentage of overhead lines, the exposure to flood and storm risks etc. and is likely to be reflected in different valuations for electricity services.

Ofgem's workhorse model is the MNL model. Heterogeneity is (in some cases) introduced by segmented estimations (e.g. by income groups or DNO). However, while Ofgem sets company specific price caps, the assumption of homogeneous preferences within DNOs is sustained. This assumption of homogeneous preferences within DNOs is problematic. Moreover, this approach is challenged by small (sub)sample sizes that can lead to convergence and robustness issues.

Our model flexibly allows for both heterogeneity across and within DNOs without presuming any segments a priori. That is, we do not segment our estimations along socio-demographic characteristics. While we stay agnostic about the sources of heterogeneity in the data, we need to specify how the attributes and levels enter the utility function. We test two linear in parameter specifications. They are discussed below. In all considered specifications we distinguish WTP from WTA to test whether the valuation of an improvement of a service differs significantly from the valuation of an equivalent deterioration. We argue that there is reason to suspect an asymmetry in the WTP and WTA, because most DNOs have a very high reliability and resilience: with the status quo reliability of 99.99 percent a customer experiences a power cut every two years. Customers might thus only be willing to pay a fairly small amount for further improvement, while the compensation they require for a deterioration of reliability (i.e. the WTA) is likely higher. This could reflect the behavioural economic concept of loss aversion for example. We de facto consider improvement and deterioration as different attributes.

In a first specification (Dummy Model) we consider the valuations for different levels of the attributes separately, working with a large set of dummy variables. As an example, in the Dummy Model the WTP for a 10 percent reduction of customers affected by major storms is measured by a separate parameter than the WTP for a 20 percent reduction of customers affected by major storms. Column 1 in table 4 lists the estimates resulting from a MNL estimation of the Dummy Model. The vector of attribute variables is $\mathbf{v} = (ASCA, ASCB, D_{0\%ug}, D_{3\%ug}, D_{5\%ug}, D_{-20\%sd}, D_{-10\%sd}, D_{+10\%sd}, D_{+20\%sd}, ...)$. We omit the indices for simplicity. Two alternative specific constants are included to control for a suspected preference of customers for the status quo alternative. One ASC dummy indicates option A, the other option B. A positive coefficient of the alternative specific constant A (ASC_A) for example demonstrates the underlying preference for choosing alternative A over and above any value that is associated with the service attributes. A negative coefficient does indicate a status quo effect. With 1.5 percent being the status quo of overhead line conversion to underground lines per annum. The dummy $D_{0\%ug}$ indicates that the offered alternative involves no undergrounding (1.5 percent less than to date). We expect a negative WTP, i.e. a WTA, for such reduced undergrounding activity. The dummy $D_{3\%ug}$ measures improvements in undergrounding activity from 1.5 to 3 percent. The dummy $D_{5\%ug}$ separately indicates an improvement from 1.5 to 5 percent. The dummy $D_{-20\%sd}$ indicates if the number of customers affected by storm damages due to black outs is reduced by 20 percent and the dummy $D_{-10\%sd}$ indicates the respective 10 percent improvement of resilience. The dummies $D_{10\%sd}$ and $D_{20\%sd}$ indicate the respective deterioration of resilience and so forth. We would expect a positive coefficient for all dummy variables that indicate improved resilience and negative coefficients for all dummies indicating deterioration. Incorporating separate dummy variables for all different service levels becomes rather a burden, if there is little variation in the respective levels. Moreover, since the valuations of the different levels of any attribute are likely to be correlated, correct model specification is likely to require the estimation of the respective covariances, implying large variance-covariance matrices and a loss of degrees of freedom.

A straight forward way of simplification is linearisation of the levels that implies a constant marginal utility from incremental service improvements. In a second specification (Linearised Model) we therefore partly linearise the model, e.g. we assume that the WTP for resilience to storms is linear in the percentage changes. This implies, for example, that the customers' valuation of a 10 percent improvement is half as high as their valuation of a 20 percent improvement. The observed utility of customer i for option j for the linearised model is specified as:

$$\begin{aligned} V_{jit} = \alpha_{pi}\mathrm{p_{jt}} + \omega_{ASCji} + \omega_{1i}\mathrm{lessundergd_{jt}} + \omega_{2i}\mathrm{moreundergd_{jt}} \\ +\omega_{3i}\mathrm{lessstorms_{jt}} + \omega_{4i}\mathrm{morestorms_{jt}} + \omega_{5i}\mathrm{lessfloods_{jt}} + \omega_{6i}\mathrm{lessCO2_{jt}} \end{aligned} \tag{14}$$

Where α_{pi} and $\omega_{\mathbf{ji}}$ are parameters to be estimated from the data and $\mathbf{v}_{jt} = (\mathrm{ASC_{jt}}, \mathrm{lessundergd_{jt}}, \mathrm{moreundergd_{jt}}, \mathrm{lessstorms_{jt}}, \mathrm{morestorms_{jt}}, \mathrm{lessfloods_{jt}}, \mathrm{lessCO2_{jt}})$ is a vector of alternative specific attributes. As in the Dummy Model, two alternative specific constants are included to control for a suspected preference of customers for the status quo alternative. With 1.5 percent being the status quo of overhead line conversion to underground lines per annum, the variable *lessundergd* indicates if the offered alternative involves no undergrounding (1.5 percent less than to date). The variable *moreundergd* is a categorical variable that measures improvements in undergrounding activity to 3 percent and 5 percent. Given the linearisation, the specification imposes the same coefficient for the 1.5 percent improvement as on the subsequent 2 percent increase in undergrounding activity. The variable *lessstorms* is a categorical variable that measures the reduction of storm damages (hence improvements of resilience to storm). It indicates whether 10 percent or 20 percent less customers are affected by blackouts due to major storms. The variable *morestorms* indicates the respective deterioration of resilience. The variable *lessfloods* is a categorical variable that indicates the number of major electricity sites exposed to flood risk (1,000 being the status quo). Finally the variable *lessCO2* indicates the percentage of current equipment and vehicles that is replaced with those using less polluting fuels. We expect a positive coefficient for all variables that indicate improved resilience and negative coefficients for all dummies indicating deterioration. The assumption of linearity in levels can lead to confounded estimated means, if the WTP does not increase linearly with the attribute level.

We test the two models based on a simple MNL estimation with the aim to choose one preferred specification. We thereby face a trade-off: on the one hand the Dummy Model does not constrain the valuations of the levels of an attribute to be linearly or otherwise related and can hence capture the preference heterogeneity more generally. On the other hand it is computationally more burdensome to relax the parameter restrictions and allow for random and possibly correlated parameters. The more parsimonious linearised model, on the other hand, has less coefficients to estimate than the Dummy Model and avoids by construction the estimation of preference correlations between the levels of any attribute. Only correlations across attributes need to be considered, which reduces the number of parameters to estimate significantly.

Table 4 lists the results of the MNL estimation for the two specifications. We report cluster robust standard errors that take the panel nature of the data (i.e. potential clustering on the respondent level) into account. The Dummy Model provides insights in how far the linearisation of valuations for different attribute levels is justified. For undergrounding, the marginal WTP seems to be decreasing in the level of undergrounding, while the marginal valuations of improved storm resilience seem to be constant (i.e. customers indeed value a 20 percent improvement twice as much as a 10 percent improvement). Model fit as measured by LL as well as by AIC and BIC is better when controlling for each attribute level separately, as it is the case in the Dummy Model.[7] However, due to the relative ease of implementation our preferred specification is the Linearised Model. This specification implies few additional parameters to estimate, is less computationally intensive than the Dummy Model and by construction less prone to correlation of the coefficients.

The Linearised Model assumes a utility that is linear in the attribute levels, but distinguishes WTP from WTA. In the following sections, all estimations are based on this linear specification. However, we

[7] As we calculate cluster robust standard errors, likelihood ratio (LR) tests are not meaningful. However, we also estimate the models without robust standard errors and perform the LR test based on these. The Dummy Model, i.e. the unrestricted model in which the linearised model is nested, is supported by the LR tests.

Table 4: Multinomial Logit Specifications

	(1) Dummy	(2) Linearised
Dummy no undergd	-0.349***	
	(0.0464)	
Dummy 3% undergd	0.270***	
	(0.0456)	
Dummy 5% undergd	0.362***	
	(0.0461)	
Dummy 20% more storm damage	-0.531***	
	(0.0524)	
Dummy 10% more storm damage	-0.364***	
	(0.0505)	
Dummy -10% less storm damage	0.184***	
	(0.0484)	
Dummy -20% less storm damage	0.313***	
	(0.0496)	
Dummy less flood risk (-50 sites)	0.229***	
	(0.0440)	
Dummy less flood risk (-100 sites)	0.162***	
	(0.0432)	
Dummy less flood risk (-150 sites)	0.246***	
	(0.0451)	
Dummy 5% less CO2	0.561***	
	(0.0415)	
Dummy 10% less CO2	0.680***	
	(0.0435)	
Less undergd ($\Delta = -1.5\%$)		-0.369***
		(0.0447)
More undergd ($\Delta = +1.5\%$)		0.180***
		(0.0228)
More storm damage ($\Delta = +10\%$)		-0.267***
		(0.0251)
Less storm damage ($\Delta = -10\%$)		0.170***
		(0.0237)
Less flood risk		0.0639***
		(0.0140)
Less CO2		0.327***
		(0.0211)
Change in bill	-0.0859***	-0.0853***
	(0.00309)	(0.00309)
ASC A	-0.766***	-0.628***
	(0.0670)	(0.0585)
ASC B	-0.908***	-0.771***
	(0.0671)	(0.0586)
Observations	36036	36036
LL	-10757.913	-10791.898
AIC	21545.8	21601.8
BIC	21673.2	21678.2

Cluster robust standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

test models that differ in the manner in which unobserved heterogeneity is accommodated: depending on the model we allow for constant/random coefficients, homogeneous/heterogeneous scale, uncorrelated/correlated coefficients. This enables us to explore the sensitivity of the estimates to the respective assumptions. We a) estimate the population moments, b) derive the individual level valuations (their means and variances) and c) examine these posterior valuations by DNO and other population segments.

5.2 Flexible Mixed Logit in WTP Space

For the official report Ofgem conducted the estimations partly on the whole sample and partly segmented by DNO. Based on the examination of cross-tables, the data was split further where significant differences in valuations were suspected. A problem with this segmented estimation is the small sample size, which hampers robust estimation, especially when estimating more complex models such as MXL or GMNL. The price coefficient might be insignificant and differential price sensitivity (e.g. by income) might not be identifiable when estimating the models for each DNO separately. Aggregation of the data can therefore be particularly important for the estimation of sensitivity to price. Since only about 150 customers were observed per DNO, we perform pooled estimation across all customers and DNOs.

To obtain baseline estimates that are comparable to those of Ofgem, we first estimate a multinomial logit model (MNL) in preference space (M1). The WTP values are derived as ratio of any attribute coefficient ω_k to the price coefficient α. Table 5 lists the estimated valuations from this baseline model. They are very close to those reported by Ofgem. As an example, while our baseline model suggests a WTP of £2.11 for an increase in undergrounding by 1.5 percent per annum over the five year price control, Ofgem's research indicated a WTP of £2.29 per annum. Under our linearity assumption, we also get a very similar value for the increase in undergrounding activity from 1.5 to 5 percent per year: while our MNL estimate suggests a WTP of £4.22, Ofgem estimates the value to be £4.36. When estimating the MNL model in WTP space (M2) the estimated WTPs correspond to those in M1. This equivalence between the models in preference and WTP space holds because all coefficients are constant.[8] The MNL model is easy to implement, but due to its restrictive nature likely to be misspecified. A mistakenly omitted scale factor for example implies that the heteroscedastic error variance is not accounted for. In the logistic choice models considered here, this leads to biased estimates in addition to wrong standard errors.

Table 5: WTP (£per annum) for attribute k derived from MNL in preference space (M1) as ratio $\frac{\omega_k}{\alpha}$

	1.5% less undergd	1.5% more undergd	10% less storm dmg	10% more storm dmg	less flood risk	less CO2
WTP (£p.a.)	**-4.32**	**2.11**	**1.99**	**-3.13**	**0.75**	**3.83**
lower limit	-5.37	1.58	1.44	-3.72	0.43	3.31
upper limit	-3.26	2.65	2.54	-2.54	1.07	4.35

The table lists the WTP estimates (and estimated standard deviations) as resulting from the baseline MNL model, which assumes that all customers have the same WTP and don't differ in their randomness of choice. The values correspond closely to those of Accent (2008).

As argued in section 3, random parameter models with flexible mixing distributions can have sig-

[8]The price coefficient in M2 can be retrieved from the constant as $-exp(-2.461) = -0.085$.

nificant advantages over more assumption-bound specifications. More flexible models are often more difficult to estimate, though, with convergence issues and problems with parameter significance (Hess, 2010). When choosing the appropriate model, researchers have to trade-off the costs and benefits of flexible mixing distributions.

We test a range of random parameter models, among them a heterogeneous scale mixed logit model (exploiting the GMNL specification) that can accommodate a random scale parameter. We clearly distinguish the ability to accommodate a random scale parameter in the model from the ability to identify this parameter. Compared to the MNL model our main model specifications relax the restriction on the scale parameter (M3, the SMNL model), allow for random preference parameters (M4, the MXL model) and a combination of both (M5, the heterogeneous scale mixed logit or GMNL model), respectively. We abstract from models that additionally allow for coefficient correlations which go beyond the correlation implied by the scale parameter.

We consider different distributional assumptions for the valuations. Due to their unboundedness to at least one side, both normal and log-normal distributed model parameters can result in unreasonably large estimates for some share of decision makers.[9] And both types of distributions come with further drawbacks: while the normal distribution may give rise to the wrong parameter sign, at least for some respondents (see Ghosh et al., 2013), the log-normal distribution has zero probability mass at zero. The log-normal assumption is therefore not suitable for modelling situations where a section of the population is indifferent to an attribute (see Rigby and Burton, 2006; Train and Sonnier, 2005). Since the attribute levels of interest are quite abstract in this experiment, some customers might find it difficult to relate to the attributes (e.g. to electricity sites exposed to flood risks or the number of customers affected by storms). We thus expect some of the customers to be indifferent and hence avoid the log-normal distribution and its inability to capture indifference for the valuation parameters. Moreover, log-normal distributions also frequently come along with estimation challenges and lastly, we do not want to pre-impose a parameter sign in cases in which it is unambiguous whether people are WTP or require compensation for the respective attribute. Undergrounding of overhead lines, for example, might be of value to some, but not to other customers. For these reasons we assume normal rather than log-nornal distributions for all WTPs and WTAs. However, we assume a log-normally distributed scale parameter ($\sigma_i \sim logN$) and the distribution of the price coefficient in WTP space depends on the model specification ($\lambda_i \sim logN$ or $\lambda_i = (\sigma_i \times \alpha_i) \sim (logN \times logN)$).

Moreover, for the reasons outlined in section 3, all models are estimated in WTP space. We rely on the GMNL framework in WTP space, imposing the respective constraints as listed in table 1. The GMNL model is estimated by simulated maximum likelihood (SML), which maximizes the marginal likelihood function after integrating over heterogeneity distributions. The pooled estimation imposes the restriction of constant tastes across replications for the same respondents. We estimate the model with different starting values to ensure that the global maximum has been reached.

Since the more flexible models come at the cost of increased computational burden, our estimation

[9] The long tail of the log-normal distribution for example often generates a large range of implausible WTP estimates.

strategy trades-off the benefits of model flexibility (such as better model fit and more realistic WTP distributions) against the cost (such as convergence issues and computation time). To achieve convergence of the random parameter models, we set $\gamma = 1$, which gives further reason to not claim identification of scale in the traditional sense. $\gamma = 1$ reflects the prior that the individual level parameters follow a mixture of normals with different means but equal variances rather than variances that are scaled proportionally.

Table 6: Homogeneous Preference Parameter Models (M1-M3)

GMNL specification	M1 MNL-pref. space	M2 MNL-WTP space	M3 SMNL-WTP space
Change in bill	-0.0853***		
	(0.00309)		
ASC A	-0.628***	-7.358***	-2.969***
	(0.0585)	(0.660)	(0.437)
ASC B	-0.771***	-9.038***	-3.581***
	(0.0586)	(0.667)	(0.443)
Less undergd (-1.5%)	-0.369***	-4.318***	-2.016***
	(0.0447)	(0.512)	(0.329)
More undergd (+1.5%)	0.180***	2.111***	0.801***
	(0.0228)	(0.262)	(0.167)
Less storm damage (+10%)	0.170***	1.992***	0.808***
	(0.0237)	(0.270)	(0.188)
More storm damage (-10%)	-0.267***	-3.128***	-2.286***
	(0.0251)	(0.287)	(0.199)
Less flood risk	0.0639***	0.749***	0.303**
	(0.0140)	(0.166)	(0.104)
Less CO2	0.327***	3.830***	1.525***
	(0.0211)	(0.233)	(0.154)
[Het] const		-2.461***	-1.616***
		(0.0197)	(0.0550)
τ		0	1.366***
		(.)	(0.0511)
Observations	36036	36036	36036
LL	-10791.9	-10791.9	-10221.44
AIC	21601.8	21601.8	20462.9
BIC	21678.2	21678.2	20547.8

Cluster robust standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6 lists the results of models with homogeneous preference parameters. M3 allows for scale heterogeneity. All estimated mean valuations have the expected and intuitive signs and significance. The significant scale parameter τ suggests that models M1 and M2 are misspecified and the means hence confounded as these models do not take the heteroscedastic error variance into account. Model fit is improved when increasing model flexibility via the introduction of the random scale parameter.

Table 7: Random Parameter Models (M4, M5)

	M4	M5
GMNL specification (WTP space)	MXL	GMNL-I
Mean		
Change in Bill	1	1
ASC A	-4.691***	-1.860***
	(0.601)	(0.354)
ASC B	-6.401***	-2.444***
	(0.605)	(0.350)
Less undergd ($\Delta = -1.5\%$)	-3.545***	-1.941***
	(0.500)	(0.290)
More undergd ($\Delta = +1.5\%$)	1.962***	0.765***
	(0.276)	(0.161)
Less storm damage ($\Delta = -10\%$)	1.777***	0.788***
	(0.277)	(0.174)
More storm damage ($\Delta = +10\%$)	-3.506***	-2.052***
	(0.300)	(0.175)
Less flood risk ($\Delta = 50$ sites)	0.493**	0.295**
	(0.172)	(0.0948)
Less CO2 ($\Delta = 5\%$)	3.172***	1.340***
	(0.276)	(0.148)
[Het] const	-2.004***	-1.019***
	(0.0237)	(0.0551)
τ	0	1.343***
		(0.0478)
SD		
ASC A	0.884***	1.101***
	(0.0729)	(0.0895)
ASC B	0.831***	0.888***
	(0.0800)	(0.107)
Less undergd ($\Delta = -1.5\%$)	1.130***	1.448***
	(0.0772)	(0.0946)
More undergd ($\Delta = +1.5\%$)	0.670***	0.875***
	(0.0578)	(0.0682)
Less storm damage ($\Delta = -10\%$)	0.613***	0.833***
	(0.0569)	(0.0697)
More storm damage ($\Delta = +10\%$)	0.610***	0.793***
	(0.0623)	(0.0802)
Less flood risk ($\Delta = 50$ sites)	0.390***	0.467***
	(0.0350)	(0.0436)
Less CO2 ($\Delta = 5\%$)	1.000***	1.249***
	(0.0447)	(0.0591)
Observations	36036	36036
LL	-9758.9	-9136.9
AIC	19551.7	18309.9
BIC	19696.1	18462.8

Cluster robust standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7 lists the results of the models with random parameters. All estimated mean valuations have the expected and intuitive signs and significance. The significant standard deviations and scale parameter τ suggest that models M1, M2 and M3 are misspecified: the means are confounded as these models do not take preference heterogeneity and/or heteroscedastic error variance into account. Model fit is improved when allowing for more flexible distributions to model the parameter heterogeneity.

Table 6 lists the results of the MNL in preference and WTP space and of the SMNL model in WTP space. Table 7 lists the results of the models with random preference and scale parameters. We do not allow for correlated coeffcients to avoid increasing complexity and computational cost of the model. Across all models signs and significance of the estimated mean valuations are consistent and in line with expectations. Consumers have a preference for the status quo alternative, which is reflected by the significant negative coefficients of both ASCs. As indicated by the negative WTP coefficients, consumers need to be compensated in case of service deterioration such as in case of less undergrounding activity or if more households are affected by blackouts due to storms. On the other hand, consumers are willing to pay for increased undergrounding activity, for improved resilience to storms and floods and for more investment (by distributor) to current mobile generation equipment and vehicles to reduce CO2 emissions, which is reflected in positive and significant estimated coefficients. Moreover, as suspected, we find significant differences between the WTP and the WTA (in absolute terms): customers are willing to pay less for a given service improvement than they would need to be compensated for in case of the equivalent service deterioration. While the baseline model (5) yields a WTP for an additional 1.5 percent of undergrounding per annum of £2.11 for example, the compensation required for 1.5 percent less undergrounding than the status quo (i.e. the WTA) is more than twice as large, namely £4.32. The same holds for the preferences for storm damages: while customers are willing to pay £1.99 for 10 percent less customers being affected by major storms, the WTA 10 percent more customers being affected is about £3.13. This finding is consistent with some degree of loss aversion of the consumers.

Despite correct signs and significance of the estimated means across all models, the estimated valuations differ in magnitude. This results from the different assumptions imposed on the parameter heterogeneity and hence the flexibility of the models. In WTP space the WTPs are unconfounded by scale, but since the scale parameter is confounded with the price parameter, differences in the estimates across models are a consequence of the impact that the inclusion of the additional parameter has on the remaining model parameters (Hess and Stathopoulos, 2013).

Moreover, the estimated standard deviations in the random preference parameter models are all significant (see table 7). The parameter heterogeneity impacts not only the distribution, but also the estimated mean valuations: when comparing MNL (M1 and M2) with MXL (M4) for example (both with constant scale), the estimated means differ. They are lower (in absolute terms) for all attributes apart from the WTA more storm damages. The significant heterogeneity in M4 might also be caused by heterogeneous scale that the model fails to capture, but the findings support the hypothesis of substantial parameter heterogeneity in the data and render the constant preference parameter models (M1, M2, M3) misspecified.

Furthermore, models that take scale into account and thus allow for the more flexible mixing distributions result in attenuated mean estimates. As discussed in section 3, omitting the scale parameter from the model can lead to biased estimated means if there is heteroscedasticity. The results of the SMNL model in WTP space (M3) as well as of the model with random preferences and scale heterogeneity (M5, M7) suggest that the more flexible model is reasonable: the coefficient τ is highly significant. However, building on the argument made in section 3, part of the heterogeneity captured by the scale parameter in the SMNL specificaiton (M3) might equally result from heterogeneity in the preference parameters that the distribution imposed fails to capture. If the covariances of all attributes and the ASCs in our model were unconstrained, this would imply the estimation of 28 additional covariance parameters and hence loss of 28 additional degrees of freedom. As these models are computationally burdensome, we abstract from these models in this analysis.

Given the discovered parameter heterogeneity and the computational burden of models with correlated coefficients, model M5 seems superior. This is supported by the model's fit: not only is the fit improved when moving from preference to willingness to pay space, but also when including preference and scale parameters separately. We report the final log-likelihood (LL), the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) in table 8. Comparing these measures of fit for two models estimated on the same data set allows the statistical significance of new model coefficients to be assessed and to inform model selection. While adding parameters increases the absolute value of LL it can result in overfitting. Both AIC[10] and BIC introduce a penalty term for the number of parameters in the model. Models with the lower AIC or BIC have better fit among this finite set of models.

Firstly, models that allow for heterogeneous preference parameters (α_i and ω_i) improve the BIC by around 2,000 points: while the BIC resulting from the MNL models (M1 and M2) is 21,678.2, the BIC in the MXL model improves to 19,696.1 (i.e. by 1,982.1). Comparably, when comparing the models that include a separate scale parameter the BIC of 20,553.8 resulting from the SMNL model with random scale but constant preference parameter (M4) improves to a BIC of 18,462.8 for the GMNL-I model that allows for random preference parameters in addition to random scale (M5) (improvement by 2,091). We also find that, ceteris paribus, the BIC is improved by around 1,200 BIC points when the random scale parameter is incorporated. Comparing the BIC of MNL in WTP space (M2) with the BIC of SMNL (M3) and comparing the BIC of MXL (M4) to the BIC of GMNL-I (M5) illustrates how the introduction of the parameters σ_i and τ leads to a better fit even when employing fit statistics that penalise the introduction of additional variables. GMNL-I (M5) has better fit as a result of separating out scale and the fact that the distribution of $\sigma_i \times \omega_i$ (M5) is different from the distribution of c_i (M4). In summary, BIC improves the more flexible the model is: both, allowing for random preference as well as scale paramter improves model fit.

The incorporation of the scale parameter improves the model fit, despite the WTP estimates being not directly impacted by scale. This is a consequence of the more flexible mixing distributions in models that include a separate scale parameter. The flexibility of the mixed logit model increases with the introduction of random parameters, a separate scale parameter and relaxed restrictions on the covariances. The improvements in model fit obtained by using a specification that allows the attribute coefficients to be the product of two separate and independently distributed coefficients are hence likely due to the fact that this more flexible distribution better explains the behaviour in the data. Gains in fit resulting from the inclusion of σ_i result at least in part from this increased flexibility.[11]

[10] AIC $= 2k - ln(L)$ and BIC $= -2ln(L) + kln(n)$.

[11] Any improvements of fit when moving from a model in preference to willingness to pay space on the other hand can be explained by the different assumptions imposed on the coefficients. If we imposed the distributional assumptions in

Table 8: Measures of Fit

	M1	M2	M3	M4	M5
LL	-10791.9	-10791.9	-10224.4	-9758.9	-9136.9
AIC	21601.8	21601.8	20468.9	19551.7	18309.9
BIC	21678.2	21678.2	20553.8	19696.1	18462.8

The table lists LL, AIC and BIC for all estimated models as they result from models estimated with 2,000 Halton draws. It shows that more flexible models that take preference and scale heterogeneity into account fit the data better. AIC and BIC penalise the introduction of additional parameters to avoid overfitting, but still indicate better fit for the more flexible models. Incorporating a scale parameter has a particularly prominent impact on model fit.

The differences in fit between models arise in the ease in which given distributional shapes can be accommodated, with advantages for different specifications in different settings (Hess and Rose, 2012). In models in which the distributional assumptions are imposed such that the models have the same flexibility, model fit should be the same. As an example, M4 and M5 would be formally equivalent if $c_i = (\sigma_i \times \omega_i)$ and both σ_i and ω_i were assumed to be distributed such that the product followed the same distribution as c_i. Model fit would be the same. We hence conclude that scale matters and should be included in the model to accommodate flexible distributions, but that the scale parameter cannot be separately identified. In preference space all mean coefficient estimates will be confounded by scale and in WTP space the price coefficient will be confounded.

A further criterion to evaluate the models, is the probability of sign reversal. For any attribute k, with $\omega_{ki} \sim N(\mu_{\omega_k}, \sigma^2_{\omega_k})$ standardisation yields $z_{\omega_{ki}} = \dfrac{\omega_{ki} - \mu_{\omega_k}}{\sigma_{\omega_k}} \sim N(0,1)$. Thus, the probability of a valuation being negative can be calculated as given in equation 15. We find that a disadvantage of model specifications incorporating scale is that the probabilities of sign reversal are slightly higher than in models without random scale (see table 9). Alternative distributions imposed on the coefficients (e.g. skewed distributions or distributions with positive or negative support) placed on the coefficients might reduce these issues of sign-reversals.

$$P(\omega_{ki} < 0) = P(z_{\omega_{ki}} < \frac{-\hat{\mu}_{\omega_k}}{\hat{\sigma}_{\omega k}}) \,\forall\, k \tag{15}$$

preference space such that the ratio of the attribute to the cost coefficient follows the same distribution as the valuations in WTP space, the same model fit would result in preference and WTP space. This is demonstrated in the most simple form with the equivalence of model fit of MNL in preference and WTP space (M1 and M2, respectively).

Table 9: Probabilities of Sign Reversal

	M4	M5
GMNL specification	MXL	GMNL-I
ASC A	0	0.046
ASC B	0	0.003
Less undergd (-1.5%)	0.001	0.09
More undergd (+1.5%)	0.002	0.191
More storm damage (+10%)	0	0.005
Less storm damage (-10%)	0.002	0
Less flood risk	0.103	0.264
Less CO2	0	0.142

Table 8 lists the probabilities of unexpected sign as resulting from the random parameter models M4 and M5. The model that allows for preference and scale heterogeneity (M5) results in the highest probability of sign reversal. While fit was best for M5, the higher probability of of sign reversal can be seen as a disadvantage of the more flexible model.

5.3 A Priori Assessment of Preference Heterogeneity

The estimated models revealed significant heterogeneity in valuations. However, all random parameter models abstract from the sources of heterogeneity, but capture it entirely in the random parameters in the model. To shed light on the drivers of the revealed heterogeneity, we also test models with interactions of attribute and respondent characteristics. Major challenges and drawbacks of such models are the difficulty to select the appropriate interactions a priori and the increased complexity of the model as the number of included variables increases. However, simple MNL models with interaction terms of attribute and respondent characteristics can provide us with first insights into the drivers of heterogeneity.

We test the interactions of income with all attributes. Of particular interest are the interactions of the cost variable with the income variable. These can reveal whether significant income effects are present. We find significant positive coefficients of the cost-income interactions. These imply that higher income consumers' utility is less senstitive to price than lower income consumers. Since the WTP for different attributes are derived as ratio of attribute coefficient to the cost coefficient, a heterogeneous cost sensitivity implies heterogeneous valuations across income groups. A lower price sensitivity for higher income consumers ceteris paribus implies they have a larger WTP and lower WTA. Moreover, the two significant positive coefficients of the CO2-income interactions imply that higher income consumers experience a higher sensitivity to carbon reductions than lower income consumers. However, albeit significant, the differences in valuations are very small. Table 10 lists the results of the MNL model with the cost-income interactions, illustrating the heterogeneity in cost sensitivities.

We also test MNL models with interactions of the attributes with the respondents' job, age, power cut experience and fuel poverty category and select those interactions that are significant. The results of the specification with likely relevant interactions are listed in table H.3 in the Appendix.

Table 10: MNL with price income interactions

	MNL
Cost (£)	-0.129***
	(0.00964)
ASC A	-0.633***
	(0.0592)
ASC B	-0.772***
	(0.0592)
Less undergd	-0.375***
	(0.0449)
More undergd	0.181***
	(0.0230)
More storm damage (+10%)	-0.272***
	(0.0253)
Less storm damage (-10%)	0.169***
	(0.0238)
Less flood risk	0.0670***
	(0.0140)
Less CO2	0.334***
	(0.0213)
CostXinc2 (£10k - £20k)	0.0348**
	(0.0114)
CostXinc3 (£21k - £30k)	0.0465***
	(0.0127)
CostXinc4 (£31k - £40k)	0.0589***
	(0.0118)
CostXinc5 (£41k - £50k)	0.0768***
	(0.0120)
CostXinc6 (£51k - £60k)	0.0870***
	(0.0145)
CostXinc7 (> £60k)	0.109***
	(0.0126)
CostXinc8 (refused)	0.0253
	(0.0146)
CostXinc9 (NA)	0.0286*
	(0.0124)
N	36036

The reference category for income is the group with less than £10k p.a.
Cluster robust standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.4 Posterior Analysis & Primary Results

To estimate the DNO specific price caps, most estimations conducted for Ofgem are segmented by DNO and based on small samples of only 150 customers per DNO. The segmented estimation would be more robust with larger samples (e.g. 1,000 observations per DNO), but this would increase the cost of the respective experiments significantly. In the previous section we introduced an alternative approach to allow for heterogeneity in valuations, namely the introduction of attribute-covariate interaction terms. We also pointed out the drawbacks that come along with such an approach. To address the challenges associated with an a priori segmentation or selection of interaction terms, we now investigate heterogeneity based on segmentation post estimation.

We perform pooled estimation across all observations and segment the estimated individual conditional mean valuations post estimation. First, we investigate the properties and merits of the estimated conditional distributions along the lines of Hess (2007): how reasonable are the posterior distributions compared to the estimated distributions? What is the importance of between and within variation? How do the probabilities of unexpected signs and the preference correlations compare? Second, we employ two types of posterior analysis to investigate posterior mean differences in valuations and respondent characteristics: are there any meaningful links of the individual level conditional valuations to socio-demographic attributes or correlates of scale such as previous experience that can help to identify different segments of respondents? Third, we shed light on three individual customer profiles, exploiting the full individual level conditional distributions, and fourth, investigate preference correlations. Finally, we combine our findings to derive aggregate valuations within and across DNOs.

5.4.1 Posterior Mean Distributions

For all estimated random parameter models the means of the individual conditional means are very close to the corresponding estimated unconditional population means, i.e. $\frac{1}{NR}\sum_{i=1}^{N}\sum_{r=1}^{R} WTP_{ir} \approx \hat{\mu}$. And as expected the posterior standard deviations of the conditional means (i.e. the variation of the conditional mean valuations between individuals) are smaller than the estimated standard deviations $\hat{\sigma}$ as the latter also includes the variation around the individual conditional means. Table 11 lists the means of the individual posterior means, their respective between standard deviations, their extreme values across all DNOs as derived from the estimated means ($\hat{\mu}$) and the share of the total unconditional variance that is explained by the variation of the conditional means ($\frac{SD}{\hat{\sigma}}$) for the preferred random parameter model M5 that flexibly allows for parameter heterogeneity across and within DNOs. For comparison, the estimated means and standard deviations from section 6 are also listed.[12]

The variation of the individual level conditional mean estimates only gives a partial representation of the heterogeneity in the data. It just captures the heterogeneity between respondents, but abstracts from the variation within individuals. Since the respondents did face a finite rather than an infinite number of choices, the meaningfulness of the estimated within variation is limited. When arguing whether the conditional estimates are more or less advantageous than the unconditional ones, a nuanced consideration

[12] The results of the posterior analysis hold likewise for the random parameter model without separate scale parameter.

Table 11: Summary statistics of individual posterior distributions across all individuals and DNOs.

M5	Mean	(Between) SD	Min	Max	$\hat{\mu}$	$\hat{\sigma}$	$\frac{SD}{\hat{\sigma}}$
ASC A	-1.850	0.622	-4.780	0.858	-1.860	1.101	0.56
ASC B	-2.463	0.463	-3.948	-0.135	-2.444	0.888	0.52
Less undergd (-1.5%)	-1.980	0.882	-5.809	0.600	-1.941	1.448	0.61
More undergd (+1.5%)	0.789	0.501	-1.152	3.009	0.765	0.875	0.57
Less storm damage	0.817	0.458	-0.756	2.712	0.788	0.833	0.55
More storm damage	-2.097	0.437	-4.114	-0.362	-2.052	0.793	0.55
Less flood risk	0.297	0.270	-0.801	1.339	0.295	0.467	0.58
Less CO2	1.383	0.867	-1.993	5.083	1.340	1.249	0.69

Table 11 lists the summary statistics of the individual conditional WTP distributions. While the means of the individual posterior means lie very close to the estimated means, the between variation of the individual posterior means is smaller than the total variance.

of context and application is therefore needed. In our application the heterogeneity of the individual mean valuations (i.e. the variation between individual types) is likely to be of most interest to electricity service providers. Electricity service companies will aim to explore customer heterogeneity, but will be likely to target the heterogeneity across individual types rather than focusing on the variation within individuals. They could link the heterogeneity in individual mean valuations back to demographics and in context of economic viability analyses the individual mean WTPs could be re-weighted depending on structural changes in the population.

The consideration of within and between variation is crucial in light of previous research that finds that the use of the conditional distributions generates a lower incidence of outliers, and related a lower incidence of sign violations. Given the use of additional information $\mathbf{y}_i$, researchers argue that inference on mean parameters of the unconditional distribution is more precise when based on the conditional distributions (Revelt and Train, 2001). However, these statements are based on the consideration of variances of the conditional means that abstract from the within variation: the variance of the conditional means is by construction smaller than the unconditional variance, as the latter is equal to the variance of the conditional means PLUS the variance of the conditional distribution around the individual means. The smaller standard errors when working with the conditional means is therefore not an indicator for more reasonable estimates, but rather a result from an exclusion of the variance around the conditional mean. Whether this is an advantage or disadvantage of the use of conditional distributions depends on the context and application of interest.

We first calculate the probability of unexpected sign based on the variation of the means, i.e. only exploiting the between variation. As expected and in line with previous findings such as those of Hess (2007), the probabilities of sign reversal shrink remarkably (see table 12). This is because the standard deviations of the conditional means do not take the variation around these means into account. The model that takes scale heterogeneity into account (M5) still shows the highest probabilities of sign reversal, but they are much lower than for the estimated population distribution (see table 9).

If the between variance captures a sufficiently large share of the total variation in a coefficient, the individual conditional means have potential to be useful in distinguishing customers (Train, 2003). Train (2003) discusses the possibility to employ cluster analysis based on conditional mean estimates in order

Table 12: Posterior Probability of Sign Reversal

GMNL specification	M4 MXL	M5 GMNL-I
ASC A	0	0.015
ASC B	0	0
Less undergd (-1.5%)	0	0.012
More undergd (+1.5%)	0	0.058
More storm damage (10%)	0	0
Less storm damage (-10%)	0	0.037
Less flood risk	0.013	0.136
Less CO2 (5%)	0	0.054

Table 12 illustrates the reduced posterior probabilities of sign reversal. Even for the flexible GMNL-I model (M5) the WTP distributions are tamed.

to respecify the model towards different segmentation. In our application the conditional means capture between 50 and 60 percent of the sample level heterogeneity (see table 11), which is likely to be large enough to use this information to distinguish between respondents, e.g. in cluster analysis. We aim to explore how observed variations in WTP and WTA correlate with observed socio-economic and demographic characteristics, in particular with factors that arguable drive scale heterogeneity.

The posterior analysis also reveals intuitive correlations between valuations both within and across attributes.[13] As an example, within attributes one could expect that customers with a high WTP for improvement of resilience have a low WTA (i.e. require high compensation) for the equivalent deterioration. 'Across attributes' on the other hand, the expected sign of the correlation is not as unambiguous, but one might expect that customers with high WTP for one attribute do also value closely related or complementary attributes more. The pairwise correlations of the individual posterior mean valuations as resulting from our preferred model M5 are listed in table 13. Most have the expected sign and intuitive and meaningful interpretations. The revealed negative posterior correlation within attributes for example reflects that customers with a lower WTA less undergrounding (more negative coefficient) are expected to be willing to pay more for more undergrounding activity (more positive coefficient) and customers who demand higher compensation for more storm damages are also willing to pay more for less storm damages. The posterior mean correlations suggest a positive correlation across attributes, i.e. customers who value carbon reductions more, are also willing to pay more for increased undergrounding activity, greater resilience to floods and storms and are less willing to accept less undergrounding. Most of the correlations are hence in line with expectations. There are only few exceptions, all related to the WTA more storm damages. There is a negative correlation with the WTA less undergrounding and a positive correlation with more undergrounding. That is, people with a higher WTA more storm damages require less compensation for less undergrounding and are willing to pay less for more undergrounding. This might indicate customers' preferences for improvements with the purpose of resilience rather than just for the sake of natural beauty (which undergrounding might be seen a means for). Posterior analysis has hence potential to reveal interesting preference relations.

[13] Since the standard errors are derived from estimates, we focus on the signs of the correlations rather than significance.

Table 13: Pairwise Correlation of Individual Posterior WTPs - M5.

WTA/WTP...	less undergd	more undergd	less storm dmg	more storm dmg	less flood risk	less CO2
1.5% less undergd	1					
1.5% more undergd	–0.30*	1				
less storm dmg	-0.20*	0.05	1			
more storm dmg	-0.16*	0.08*	-0.04	1		
less flood risk	-0.21*	0.10*	0.14*	0.12*	1	
less CO2	-0.30*	0.12*	0.15*	0.15*	0.14*	1

* $p < 0.05$

Table 13 illustrates that the posterior analysis reveals meaningful correlations within and across attributes.

5.4.2 Posterior Tests of Mean Differences

Posterior analysis of the individual level conditional estimates allows to test for and explore heterogeneity of valuations without an a priori selection of the supposedly relevant interaction terms. The increased model complexity and the subjectivity implied by the a priori covariate selection can hence be avoided.

We perform two types of posterior analysis: first we test for mean differences in individual level posterior mean valuations across different covariate categories. E.g. we test whether the valuations differ across income or job categories. Second, we cluster the individual level posterior mean valuations based on a k-means clustering algorithm and test mean differences in valuations and in respondent characteristics across the clusters. E.g. we test whether the mean income differs across the clusters.

First, the posterior tests of mean differences in valuations across covariate categories reveal significant heterogeneity of valuations across categories. When testing mean differences of individual mean WTPs across income categories for example, we find that the valuations for more and less undergrounding activity, for less floods damages as well as the valuations for less CO2 vary significantly by income. Higher income categories tend to have higher valuations. Moreover, apart from the valuations for storm resilience, all valuations differ across job categories. And the mean valuations across customer groups with different power cut experience vary significantly for all attributes apart from those regarding less storm and less flood damages. These findings of the posterior analysis are consistent with the results of the MNL model with the respective covariate-attribute interactions (see section 5.3). In a simple MNL specification in preference space the coefficient of the cost-income interactions are significant and result in higher valuations for higher income consumer categories.

Second, we perform k-means clustering on the estimated individual level posterior mean valuations to partition the data into clusters, in which each observation belongs to the cluster with the nearest mean. Several numbers of clusters k were tested. Starting from $k = 2$ the number of clusters in the population was increased until significant mean differences in valuations were found. This was the case for $k = 4$. We partition the observations into clusters that take all five valuations of an individual simultaneously into account. This could be of interest when partitioning the customers into groups with similar preferences for electricity service *bundles*, as it yields clusters of respondents that are most similar in their valuations across multiple attributes and would hence be likely to choose similar service bundles. To relate the clusters of individual posterior valuations back to the customer characteristics we test for mean differences in (socio-)economic and demographic characteristics. We find that the customers differ remarkably across the valuation clusters.[14] As an example, the average electricity bill differs significantly across the valuation clusters. In fact, the mean differences in bill size, socio-economic status, income and age across the multidimensional valuation clusters are all significant. We also find significant differences with respect to the respondents' past experience with power outages. As an example, the experience with power cuts in the past three years differs significantly across the valuation clusters: respondents in the cluster with the lowest mean WTP for more and the lowest WTA of less undergrounding services have the least experience with power cuts. Respondents who have not experienced any power outages in the past seem to not value the electricity services as much as respondents who have already experienced outages. Table 14 summarises these results from the cluster analysis. *yes* indicates that the respective respondent characteristic differed significantly across the four clusters of the posterior attribute valuations. *no* indicates that no significant mean differences across the clusters were identified.

Table 14: Mean Differences across Posterior Valuation Clusters

	Valuation Clusters ($k = 4$)
Bill size	yes
SEG	yes
Age	yes
Income	yes
Fuel Poverty	no
Power cut exp.	yes

For each attribute the table indicates which socio-economic and demographic characteristics differ significantly across the estimated individual posterior valuation clusters. *yes* indicates that the respective respondent characteristic differed significantly across the four clusters of the posterior attribute valuations. *no* indicates that no significant mean differences across the clusters were idenified.

Such cluster analysis that relates posterior valuations back to respondent characteristics can potentially be used to respecify the model. One possibility would be to estimate the preferred heterogeneous scale mixed logit specification (here M5) separately on the groups identified through the clustering.

Both types of posterior analysis (testing mean differences in valuations across covariate categories or testing mean differences of valuations and respondent characteristics across multidimensional valuation clusters) allow to flexibly test for heterogeneity of valuations, without an a priori selection of the supposedly relevant interaction terms and without increasing model complexity to an unmanageable degree.

5.4.3 Three Customer Profiles

Useful marketing information can be obtained by examining the conditional estimates for each individual. For this analysis we also estimate the individual posterior standard deviations that are a measure of the variation around the individual posterior mean. We exemplarily present three customer profiles to illustrate the potential of posterior analysis for consumer specific marketing and provision of services. Table

[14] This is done using Stata's *mvtest means* command which performs one-sample and multiple-sample multivariate tests on means. These tests assume multivariate normality.

15 lists the posterior means and standard deviations for the three customers along with the probability of opposite coefficient sign for each valuation and the posterior population means for each attribute.

Customer 1 has a particular high WTP for more undergrounding activity. The respondent also asks for compensation for less undergrounding, but in contrast to the average customer, his valuations are not as asymmetric. Only when it comes to storm resilience he asks for a relatively high compensation for deterioration of the resilience to storms as compared to what he is willing to pay for improved resilience. Lastly, he is willing to pay more than £2 per annum for CO2 savings. The respondent is above 50, belongs to the social middle class C1 and has an annual energy bill of £360. He lives in a rural area and is customer of DNO Northern Powergrid. He has not experienced any powercuts in the last three years.

Customer 2 is an example of an individual that is willing to pay for service improvements and asks for fairly symmetric compensation in case of deterioration. He is willing to pay £1.64 p.a. for more undergrounding and asks for almost the same amount of compensation in case of deterioration (£1.76 p.a.). He has an above average WTP for improved storm resilience (£2 p.a.), which might be due to his previous experience with power cuts. Lastly, he has a relatively high WTP for CO2 reductions. The respondent belongs to the rather deprived social class DE and is also above 50 years old. He has an annual energy bill of £400 and is customer of SSE Hydro.

The third customer lives in a rural area, is above 50 years old and belongs to the upper social class AB. He has a particularly high WTP for CO2 reductions. He is willing to pay on average £4.2 p.a.. He is willing to pay about £1 for more undegrounding and slightly less for improved resilience to storms. As for the average consumer, his WTP for less flood damages is lowest.

With the posterior standard deviations and the distributional assumptions (in our case normality) we derive the probability that an individual's valuation is of unexpected sign. This exploits the within variation. For all three individuals the probability of opposite coefficient sign is close to zero for most of the considered attributes. Only for services that improve resilience to flooding it is relatively likely that the individual is in fact not willing to pay.

In summary, we demonstrate that while the posterior means lie very close to the estimated means, the posterior between standard deviations for the random coefficient models with and without scale are smaller and the ranges of the valuations consequently more bounded since the variation of the conditional means abstracts from the variation around the means. In the context of electricity services, this can prove useful when targeting the preference variation across rather than within consumers. When working with the posterior distribution there is consequently a lower incidence of sign violations. While the higher probability of sign reversal appeared to be a disadvantage of the estimated distributions, posterior analysis can tame the distribution of WTPs, even when an already flexible GMNL model with scale is estimated in WTP space.

5.4.4 Posterior Aggregate Valuations

Given the merits of the conditional distributions, we exploit them to segment the data and derive summary statistics as well as the total valuations of the different attributes by DNO. Exemplary summary statistics

Table 15: Three Individual Customer Profiles

Customer Profiles	1	2	3
E_i(ASC2)	-2.09	-1.37	-0.61
SD_i(ASC2)	0.40	0.92	0.51
P(sign reversal)	0%	7%	12%
E_i(ASC3)	0.83	0.58	-0.10
SD_i(ASC3)	0.11	0.23	0.05
P(sign reversal)	0%	1%	2%
E_i(less undergd)	-2.09	-1.76	-1.65
SD_i(less undergd)	0.33	0.61	0.39
P(sign reversal)	0.00	0.00	0.00
E_i(more undergd)	2.42	1.64	1.07
SD_i(more undergd)	0.58	0.67	0.15
P(sign reversal)	0%	1%	0%
E_i(more storm dmg)	-2.06	-2.50	-1.53
SD_i(more storm dmg)	0.19	0.47	0.26
P(sign reversal)	0%	0%	0%
E_i(less storm dmg)	0.63	2.06	0.84
SD_i(less storm dmg)	0.32	0.91	0.65
P(sign reversal)	2%	1%	10%
E_i(less flood dmg)	0.10	0.27	0.18
SD_i(less flood dmg)	0.17	0.23	0.19
P(sign reversal)	27%	12%	17%
E_i(less CO2)	2.82	1.70	4.21
SD_i(less CO2)	0.44	1.26	1.34
P(sign reversal)	0%	9%	0%
DNO	9	6	9
annual bill	360	400	420
rural	yes	no	no
SEG	C2	DE	AB
age	50+	50+	50+
annual income	NA	NA	10k - 20k
powercut exp	no	yes	no

Table 15 lists the posterior mean valuations (£per customer per year), the posterior standard deviations as well as individual characteristics. The posterior standard deviations were derived using the software tool suggested by Hess (2010).

for the posterior WTP distributions for undergrounding segmented by DNO as resulting from M5 can be found in table H.4 in the Appendix. The formular for the calculation of the total valuations is given in equation 16. To refine the estimates of the DNO specific aggregate valuations of undergrounding for example, we first add up the individual posterior mean valuations of those customers with a positive WTP for increased undergrounding activity within the respective DNO in the sample and then scale these valuations proportionally to the level of the DNO. Table H.4 lists the posterior individual mean valuations by DNO exemplarily for M5. The differences in the posterior means and the posterior standard deviations confirm that the posterior valuations vary across and within DNOs. Figure 2 and 3 plot the posterior distributions of the WTP for increased undergrounding activity and improved storm resilience respectively by DNO as they result from model M5. The company names have been removed for confidentiality. The graphs illustrate the substantial heterogeneity in valuations within and across the DNOs.

Figure 2: Posterior WTP Distributions for Undergrounding by DNO

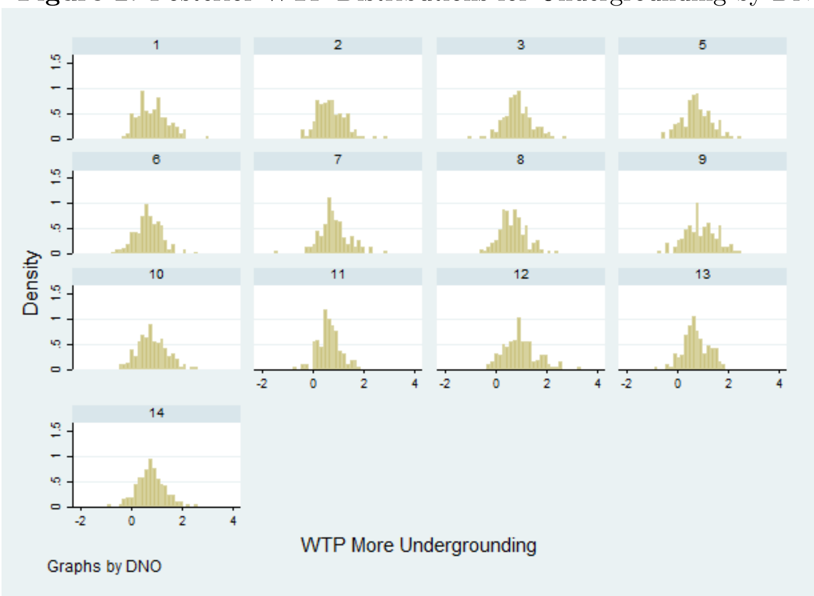


Figure 2 plots the posterior distributions of the WTP for undergrounding by DNO. There is substantial heterogeneity in valuations within the DNOs.

The total valuation within the population can then be calculated by adding up the valuations across the DNOs. $\mathbb{1}_{\mathbb{R}+}$ is an indicator function that indicates whether the individual mean WTP is positive, n_{dno} and N_{dno} indicate the number of customers per DNO within sample and population respectively.

$$\text{Total } WTP_{gmnl} = \sum_{dno=1}^{14} \underbrace{\frac{N_{dno}}{n_{dno}} * \underbrace{\sum_{i=1}^{n} \mathbb{1}_{\mathbb{R}+}[\underbrace{\overline{WTP}_{i,dno}}_{\text{indiv. post. mean WTP}}]}_{\text{within sample WTP within DNO}}}_{\text{WTP within DNO}} \tag{16}$$

A major drawback of conditional distributions lies in the use in out-of-sample forecasting. When moving from a sample to a population (as we do in this paper), it becomes necessary to associate specific conditional parameter estimates from the sample with each individual in the population. This mapping

Figure 3: Posterior WTP Distributions for Improved Storm Resilience by DNO

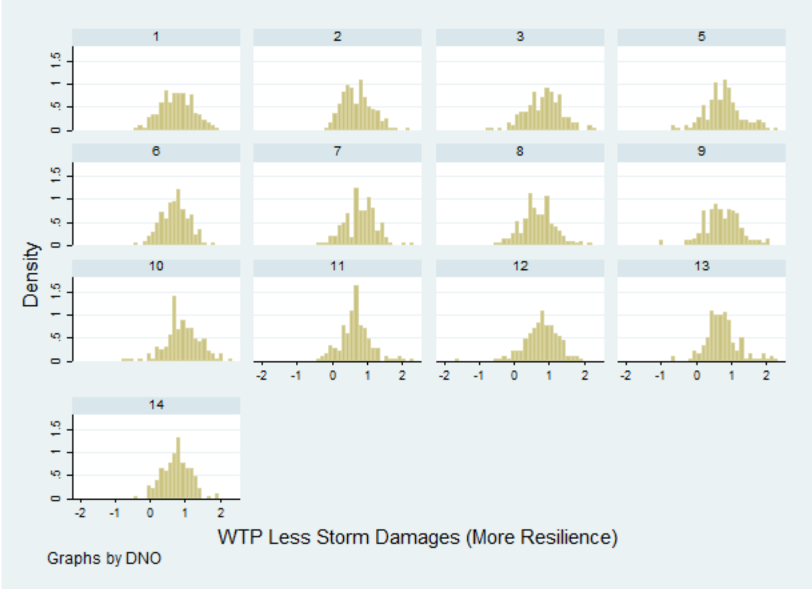


Figure 3 plots the posterior distributions of the WTP for improved resilience to storms by DNO. There is substantial heterogeneity in valuations within and across the DNOs.

of conditional parameters onto a population is a not always trivial, requiring some form of posterior analysis linking the conditional distribution to respondent specific information (e.g. income) that is also available at the population level. In our case, such information is available, but in many other cases the resepective data might be missing.

6 Policy Implications

We calculate the total attribute valuations by DNOs and across DNOs that are implied by our approach to demonstrate its practical implications for the price controls. The valuations resulting from the model currently employed by Ofgem, i.e. from the MNL model, serve as our baseline to show that the assumption of homogeneous valuations (i.e. a constant WTP) can lead to suboptimal incentives and ultimately to inefficient service levels. If there is substantial parameter heterogeneity, the estimated valuations can differ substantially depending on whether the restrictive MNL or the more flexible random parameter models are employed. In our application for example we find a significant amount of heterogeneity in customers' valuations, not only across but also within DNOs, that renders the restrictive MNL model incorrect: MNL yields unreasonably high WTP estimates and is likely to result in inefficiently high allowances and thus too much investment in undergrounding.

We argue that our more complex yet less restrictive and straightforwardly implementable estimation strategy yields more reasonable WTP estimates, fits the data better and could facilitate regulation that reflects customer preferences more realistically. While the heterogeneous scale mixed logit model (GMNL specification) in WTP space refines the mean estimates, posterior analysis can shed light on the individual WTP distribution within the population and inform the socially optimal payment plan and distribution of services. Our approach can facilitates more efficient incentive setting, both with respect to allocative and distributive efficiency.

DNOs provide services with public good characteristics (such as undergrounding) as well as services with private good characteristics (such as customer care in case of blackouts). The efficiency implications of our approach also differ depending on the type of service provided as explained in section 2.

In the case of public goods and services the correct estimation of the aggregated WTP is crucial to achieve allocative efficiency: since $n_{dno} * \overline{WTP}_{dno} = \sum_{i=1}^{n_{dno}} WTP_i$ and recalling that the estimated means are not significantly different from the posterior means (i.e. $\overline{WTP}_{dno} \approx 1/n_{dno} \sum_{i=1}^{n_{dno}} \overline{WTP}_i$), more precise estimation of the mean and hence total vertically aggregated WTP within DNOs can be decisive for whether the respective service improvement is evaluated as economically viable or not. The individual WTP distributions can then be exploited to inform the optimal payment plan. As all consumers have to consume the same quantity of the non-excludable service, the individual WTP cannot be exploited to differentiate service provision.

Investment in undergrounding of transmission and distribution networks for example is considered a public service that is provided to all customers. Undergrounding lowers the risk of damages in extreme events such as major storms or floods and reduces maintenance costs. It can thus improve reliability and safety of supply. Moreover, the visual amenity is improved (McNair et al., 2011). As McNair et al. (2011) emphasise, the household benefits are key component in the economic evaluation of undergrounding of overhead lines. The main question to evaluate its economic viability is whether the households' aggregate WTP for undergrounding exceeds the difference between the capital cost of undergrounding and the present value of ongoing cost savings for network operators. To define the DNO specific allowances, Ofgem estimated the national average customer WTP for the undergrounding of 1.5 per cent overhead lines over the course of five years (£2.29) and multiplied it by the number of customers served by each DNO. For DNO 10 for example £5 million were allowed by Ofgem following one of the customer WTP surveys. Ofgem's valuations correspond to our estimated customer valuations as derived from the vanilla MNL model: for DNO 10 our estimation implies a total WTP of £4.98 million ($2,359,391$ customers x $£2.11 = £4,978,785$) for an increase in undergrounding activity by 1.5 percent. When aggregating the DNO specific estimates across all DNOs (excluding EDF London), this implies a total WTP of about £56.6 million to increase undergrounding of overhead lines by 1.5 percent per annum (see table 16). This is very close to Ofgem's total allowance for the undergrounding of existing overhead lines that was set to £60.6 million for the five years from 2010 to 2015. However, given our analysis this estimated cap is likely to be too high.

As our results in section 6 and 7 showed, the MNL results in confounded mean estimates of the WTP. The models allowing for random parameters (flexible mixing distributions) yield different mean and hence aggregated valuations. Table 16 illustrates the practical significance of the refined estimation strategy. The table lists the total valuations per DNO (£per annum) for an increase of undergrounding activity from by 1.5 to 3 percent per annum. These aggregate values were derived from the estimated valuations per customer (£per annum) presented in tables 6 and 7. The customer level valuations were multiplied with the customer numbers per DNO as listed in table H.2 in the Appendix and resulted in the valuations

Table 16: Total WTP (£per annum per DNO) for undergrounding 1.5% more of existing overhead lines (proportionally scaled with number of customers in DNO as listed in table H.2)

1.5% more undergd	**M1/M2 MNL**	**M4 MXL**	**M5 GMNL**
1	5,516,409	5,275,323	2,311,313
2	5,163,554	4,709,555	1,795,831
3	7,421,273	7,005,532	3,023,690
5	4,712,683	4,383,379	1,794,435
6	1,563,168	1,405,633	530,340
7	6,192,551	5,780,252	2,394,334
8	4,765,683	4,306,830	1,563,004
9	3,325,011	3,184,523	1,401,679
10	4,978,785	4,803,725	2,062,617
11	2,319,812	2,042,051	722,015
12	3,252,214	3,120,650	1,461,475
13	3,133,969	2,836,783	1,131,818
14	4,205,623	3,896,866	1,572,062
Total	**56,550,735**	**52,751,100**	**21,764,613**
% of MNL base	100%	93%	38%
% of MXL base		100%	41%

Table 16 illustrates that the consideration of parameter heterogeneity can have a remarkable impact on the estimated aggregate valuations. In the case of undergrounding the allowances implied by the baseline model (in total £55.5m) would be too high. When considering preference an scale heterogeneity in the population and aggregating the individual mean WTP as suggested by equation 16, the value of undergrounding to the customers is much lower (£21.7m).

listed in table 1.6.

The valuations derived from the MNL model result in remarkable overestimation of the value of undergrounding. Moving from MNL to MXL, i.e. from homogeneous to heterogeneous preferences, reduces the estimated total valuations by up to 10 percent. When a heterogeneous scale parameter is introduced, the valuations are reduced further and can be less than half of the valuations derived from MNL. These differences in means highlight that the current manner to inform incentive rates is likely to result in too high allowances and hence overinvestment in undergrounding.[15]

A further example for a resilience service with public good characteristics is the investment in grid resilience to storms. The differences in aggregate valuations across models are comparable to the case of undergrounding as can be seen in table 17. The number of customers currently affected by storms differs remarkably by DNO. The discrete choice experiment therefore defined the changes in storm resilience relative to the DNO specific status quo. For comparison of the total valuations, we transform the estimates in valuations per 1,000 customers.

[15]The differences might be smaller in cases where customers choose less randomly. In cases where heterogeneity in the randomness of choice is likely, though, accounting for scale heterogeneity proves essential for more realistic estimates.

Table 17: Total valuations for improved resilience to storms (£per 1,000 customers per annum) by DNO

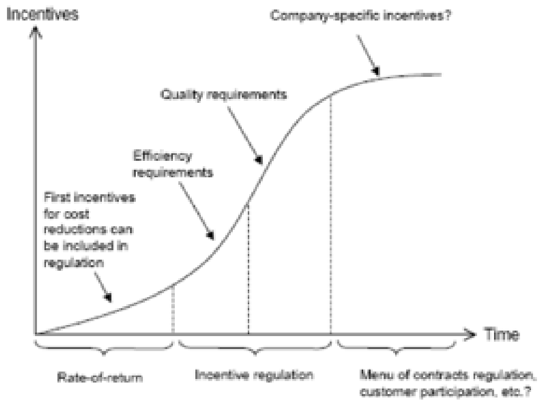
WTP less storm damage			
£per 1,000 customers	M1/M2	M4	M5
per DNO	**MNL**	**MXL**	**GMNL**
1	289,441	259,325	119,729
2	270,927	236,620	106,442
3	269,576	251,807	117,111
5	1,112,717	969,694	447,774
6	184,541	162,450	72,373
7	417,751	376,800	178,423
8	750,154	653,582	292,012
9	314,029	275,272	129,364
10	783,698	721,738	370,188
11	365,156	312,980	129,831
12	219,395	200,509	98,020
13	493,310	435,892	189,654
14	661,996	588,732	270,251
Total	**6,132,691**	**5,445,402**	**2,521,172**
% of MNL base	100%	89%	41%
% of MXL base		100%	46%

Table 17 illustrates that the consideration of parameter heterogeneity can have a remarkable impact on the estimated aggregate valuations. In the case of storm resilience the valuation implied by the baseline model is £0.6m per annum per 1,000 customers less affected by storms. When considering preference an scale heterogeneity in the population and aggregating the individual mean WTP as suggested by equation 16, the value of storm resilience as measured per 1,000 customers affected by storms is much lower (£0.2m per 1,000 customers per annum).

GMNL could thus improve the estimate of the optimal level of resilience services and posterior analysis could help to determine the optimal payment plan. With smart grid technologies, however, storm resilience is likely to become more rivalrous and excludable, as remote control and battery technologies on a local grid level will play an increasingly important role. Moreover, small-scale energy production units might be located closer to users, i.e. residential areas, leading to changes in the everyday landscape. Being natural monopolies, the DNOs have little incentive to pay attention to consumer preferences or to engage the public, unless incentivised to do so. RIIO-ED1 does provide such incentives as it explicitly rewards consideration of customer preferences. Political sensitivities regarding differentiated service quality provision are likely, but compared to the central government, local governments may be more inclined to assign differentiated quality services. This implies that with the development of smart-grids, differentiated service quality levels across the country and within DNOs could result. As an example, local network operators could differentiate their electricity service quality in terms of reliability or customer service during black-outs. Consumers who are willing to accept a lower degree of reliability or who are willing to give local grid operators access to their electric appliances to provide demand response, for example, could be charged lower prices than consumers who are willing to pay for a high degree of supply reliablity, a reliable customer service and/or do not want to provide demand response.

We recommend that, where economically and technically feasible, companies should be rewarded for providing more targeted, customer specific services. One scenario that could be considered for example is to allow for and reward differentiated, customer specific distribution charges. Price discrimination

Figure 4: The Evolution Patterns of Regulation



Source: Viljainen (2005), p.33

could ensure revenues that cover the companies' fixed costs (i.e. result in sufficient return on investment) and lead to a more efficient distribution of services in the population. Vulnerable consumers could be protected by a separate separate price cap. Figure 4 illustrates the potential transition to company specific incentives in which customer specific contracting may be taken into account and rewarded.

For such incentive regulation posterior analysis as suggested in this paper could prove particularly useful, as it gives insight into the distribution of preferences within the population (rather than just the mean or the aggregate). The more excludable and differentiated the resilience services become, the more the individual WTP distributions has potential merits for both, allocative as well as distributive efficiency.

7 Conclusion

We present an alternative approach to explore customer preferences and WTP for resilience of the electricity grid and demonstrate its practical relevance using the example of the UK's incentive regulation scheme for distribution network companies. Our analysis is based on data from the stated preference survey 'Expectations of DNOs & Willingness to Pay for Improvements in Service' that was conducted for Ofgem to inform the price control period from 2010 to 2015. Valuations of the undergrounding of overhead lines and the grid resilience to storms are in focus. In contrast to the existing manner to inform incentive rates, we flexibly account for heterogeneity in customers' valuations and choice behaviour, across and within the DNOs. More precisely, we estimate the heterogeneous scale mixed logit (GMNL) model in WTP space and perform posterior analysis to further improve the estimates. A main advantage of estimation in WTP space is that the distributional assumptions are directly imposed on the WTP.

We segment the data only post estimation to derive the summary statistics and the total valuations per DNO. The aim is to derive the incentives as they would result for each DNO based on the posterior estimates. Our results are consistent with previous literature that suggests that models which accommodate scale heterogeneity or are estimated in WTP space yield more realistic WTP distributions in the sense that they are more bounded and suffer less from outliers.

Our estimates of interest are all significant and have the expected signs: while customers ask for compensation for deterioration of resilience, they are willing to pay for services that improve resilience such as the undergrounding of overhead lines or the reduction of the number of customers that is affected by blackouts due to storms. We also find significant differences between WTP and WTA for a given service attribute. Our results suggest significant parameter heterogeneity and model fit is best for models that allow for heterogeneous preferences and scale. Our approach delivers more reasonable estimates of customers' valuation of grid resilience, not only on the individual, but also on the aggregated level of the DNOs and the population. We show that the assumptions on heterogeneity in the population can have significant implications for the price controls. If the valuation of resilience services is based on the current methodology, an inefficient supply of service quality is likely. We specifically highlight that the current manner to inform incentive rates, which assumes homogeneous valuations in the population, is likely to result in too high allowances and hence overinvestment in undergrounding.

We agree with Ofgem that incentives should be tailored for the DNOs individually, but recommend to also take heterogeneity within the customer base of any DNO into account. We suggest that, where economically and technically feasible, companies should be rewarded for providing more targeted, customer specific services. For such incentive regulation posterior analysis as suggested in this paper could prove particularly useful, as it gives insight into the distribution of preferences within the population (rather than just the mean or the aggregate). The more excludable and differentiated the resilience services become, the more the exploitation of the individual WTP distributions can matter for both, allocative as well as distributive efficiency. Our approach can deliver more realistic estimates, is straightforward to implement and could foster more targeted incentive setting.

References

Accent (2008): "Expectations of DNOs & Willingness to Pay for Improvements in Service," Report on research undertaken for customers wtp for improved services, Ofgem, `https://www.ofgem.gov.uk/ofgem-publications/47387/1704rep04final.pdf`.

Akcura, E. and M. Weeks (2014): "Valuing Reliability in the Electricity Market," *Working Paper, Faculty of Economics, University of Cambridge, UK.*

Breidert, C., M. Hahsler, and T. Reutterer (2001): "A REVIEW OF METHODS FOR MEASURING WILLINGNESS-TO-PAY," `http://www.researchgate.net/publication/242382759_A_Review_of_Methods_for_Measuring_Willingness-to-Pay`.

Daly, A., S. Hess, and K. Train (2012): "Assuring finite moments for willingness to pay in random coefficient models," *Transportation*, 39, 19–31, `http://dx.doi.org/10.1007/s11116-011-9331-3`.

Devine-Wright, P., H. Devine-Wright, and F. Sherry-Brennan (2010): "Visible technologies, invisible organisations: An empirical study of public beliefs about electricity supply networks," *Energy Policy*, 38, 4127 – 4134.

Electricity North West (2008): "Network Investment Plan 2010 to 2015,",, `http://www.enwl.co.uk/docs/about-us/network-investment-plan.pdf`.

Energy Network Association (2014): "Electricity Industry Review 18," Tech. rep., ENA.

Fiebig, D. G., M. P. Keane, J. Louviere, and N. Wasi (2010): "The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity." *Marketing Science*, 393–421, `http://dblp.uni-trier.de/db/journals/mktsci/mktsci29.html#FiebigKLW10`.

Fosgerau, M. (2006): "Investigating the distribution of the value of travel time savings," *Transportation Research Part B: Methodological*, 40, 688 – 707.

——— (2007): "Using nonparametrics to specify a model to measure the value of travel time," `https://ideas.repec.org/p/pra/mprapa/12009.html`.

——— (2014): "Nonparametric approaches to describing heterogeneity," in *Handbook of Choice Modelling*, Edward Elgar Publishing, Chapters, chap. 11, 257–267, `https://ideas.repec.org/h/elg/eechap/14820_11.html`.

Garcia, P. (2005): "Measuring Willingness-To-Pay in Discrete Choice Models with Semi- Parametric Techniques," Econometrics, EconWPA, `http://EconPapers.repec.org/RePEc:wpa:wuwpem:0504007`.

Ghosh, S., B. Maitra, and S. S. Das (2013): "Effect of Distributional Assumption of Random Parameters of Mixed Logit Model on Willingness-to-Pay Values," *Procedia - Social and Behavioral Sciences*, 104, 601 – 610, 2nd Conference of Transportation Research Group of India (2nd CTRG).

Greene, W. H. and D. A. Hensher (2013): "Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model," *Applied Economics*, 45, 1897–1902, `http://ideas.repec.org/a/taf/applec/45y2013i14p1897-1902.html`.

Hensher, D. A. and W. H. Greene (2011): "Valuation of Travel Time Savings in WTP and Preference Space in the Presence of Taste and Scale Heterogeneity," *Journal of Transport Economics and Policy*, 45, 505–525, `http://ideas.repec.org/a/tpe/jtecpo/v45y2011i3p505-525.html`.

Hess, S. (2007): "Posterior analysis of random taste coefficients in air travel behaviour modelling," *Journal of Air Transport Management*, 13, 203 – 212.

——— (2010): "Conditional parameter estimates from Mixed Logit models: distributional assumptions and a free software tool," *Journal of Choice Modelling*, 3, 134 – 152.

Hess, S. and J. Rose (2012): "Can scale and coefficient heterogeneity be separated in random coefficients models?" *Transportation*, 39, 1225–1239, `http://dx.doi.org/10.1007/s11116-012-9394-9`.

Hess, S. and A. Stathopoulos (2013): "A mixed random utility Random regret model linking the choice of decision rule to latent character traits," *Journal of Choice Modelling*, 9, 27 – 38, issues in Choice Modelling: selected papers from the 13th International Conference on Travel Behaviour Research.

Hole, A. R. and J. R. Kolstad (2010): "Mixed logit estimation of willingness to pay distributions: a comparison of models in preference and WTP space using data from a health-related choice experiment," Working Papers in Economics 03/10, University of Bergen, Department of Economics, `http://ideas.repec.org/p/hhs/bergec/2010_003.html`.

Jamasb, T., L. Orea, and M. Pollitt (2012): "Estimating the marginal cost of quality improvements: The case of the UK electricity distribution companies," *Energy Economics*, 34, 1498–1506, `http://ideas.repec.org/a/eee/eneeco/v34y2012i5p1498-1506.html`.

Jamasb, T. and M. Pollitt (2007): "Incentive Regulation of Electricity Distribution Networks: Lessons of Experience from Britain," *Energy Policy*, 35, 6163–6187, `http://ideas.repec.org/p/cam/camdae/0709.html`.

Keane, M. and N. Wasi (2013): "Comparing Alternative Models Of Heterogeneity In Consumer Choice Behavior," *Journal of Applied Econometrics*, 28, 1018–1045, `http://ideas.repec.org/a/wly/japmet/v28y2013i6p1018-1045.html`.

Li, B. (2011): "The multinomial logit model revisited: A semi-parametric approach in discrete choice analysis," *Transportation Research Part B: Methodological*, 45, 461 – 473.

Littlechild, S. (????): "Regulation and customer engagement," .

McNair, B. J., J. Bennett, D. A. Hensher, and J. M. Rose (2011): "Households' willingness to pay for overhead-to-underground conversion of electricity distribution networks," *Energy Policy*, 39, 2560 – 2567.

National Grid (2015): "UK Profile,",, `http://investors.nationalgrid.com/about-us/our-markets/uk-profile.aspx`.

Ofgem (2014): "Network Regulation - The RIIO Model,",, `https://www.ofgem.gov.uk/network-regulation-riio-model`.

Revelt, D. and K. Train (2001): "Customer-Specific Taste Parameters and Mixed Logit: Households' Choice of Electricity Supplier," Econometrics 0012001, EconWPA, `http://ideas.repec.org/p/wpa/wuwpem/0012001.html`.

Rigby, D. and M. Burton (2006): "Modeling Disinterest and Dislike: A Bounded Bayesian Mixed Logit Model of the UK Market for GM Food," *Environmental and Resource Economics*, 33, 485–509, `http://EconPapers.repec.org/RePEc:kap:enreec:v:33:y:2006:i:4:p:485-509`.

Rose, J., S. Hess, W. Greene, and D. Hensher (2012): "The Generalised Multinomial Logit model: Misinterpreting scale and preference heterogeneity in discrete choice models or untangling the un-untanglable?" .

Saint Consulting (2009): "National Survey of Opposition to Power Transmission Lines.",, `http://tscg.biz/saintblog/2009/05/national-survey-opposition-to-power-transmission-lines.html(lastaccessed5thNovember2009)`.

Soini, K., E. Pouta, M. Salmiovirta, M. Uusitalo, and T. Kivinen (2011): "Local residents perceptions of energy landscape: the case of transmission lines," *Land Use Policy*, 28, 294 – 305.

Tobiasson, W., C. Beestermoeller, and T. Jamasb (2015): "Public Engagement in Electricity Network Development: A Case Study of the Beauly Denny Project in Scotland," *EPRG Working Paper 1506 Cambridge Working Paper in Economics*, `http://www.eprg.group.cam.ac.uk/wp-content/uploads/2015/03/EPRG-WP-1506.pdf`.

Train, K. (2003): *Discrete Choice Methods with Simulation*, no. emetr2 in Online economics textbooks, SUNY-Oswego, Department of Economics, `http://ideas.repec.org/b/oet/tbooks/emetr2.html`.

Train, K. and G. Sonnier (2005): *Applications of Simulation Methods in Environmental and Resource Economics*, Dordrecht: Springer Netherlands, chap. Mixed Logit with Bounded Distributions of Correlated Partworths, 117–134, http://dx.doi.org/10.1007/1-4020-3684-1_7.

Train, K. and M. Weeks (2004): "Discrete Choice Models in Preference Space and Willingness-to Pay Space," Cambridge Working Papers in Economics 0443, Faculty of Economics, University of Cambridge, http://ideas.repec.org/p/cam/camdae/0443.html.

Viljainen, S. (2005): "Regulation Design in the Electricity Distribution Sector - Theory and Practice," Ph.D. thesis, Lappeenranta University of Technology, http://www.lut.fi/documents/10633/138922/2005-Viljainen-Thesis.pdf/e9581d34-fbe3-4d7e-9e5d-66c5830f630e.

W.H. Greene, D.A. Hensher, J. R. (2005): "Using classical simulation-based estimators to estimate individual wtp values," in *Applications of Simulation Methods in Environmental and Resource Economicss*, ed. by A. A. E. R. Scarpa, Springer Publisher, Dordrecht, pp. 1733 Ch. 2, http://ideas.repec.org/b/oet/tbooks/emetr2.html.

Wooldridge, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*, vol. 1 of *MIT Press Books, The MIT Press*, The MIT Press, http://ideas.repec.org/b/mtp/titles/0262232197.html.

8 Appendix

RPI-X and RIIO

From 1990 to 2015 the incentive regulation model of distribution networks in Britain has been a hybrid of different schemes that incentivised operating expenditure (Opex), capital spending (Capex) and quality of service separately: under the so-called RPI-X regulation the utilities' controllable Opex was incentivised by benchmarking against an efficient frontier made up of the best practice DNOs in the sector. The allowed Opex was set such that it required DNOs to close a specific proportion of their performance gap relative to the frontier during the price control period. To assess the required level of Capex over a price control period, the utilities had to submit a business plan with the projected capital expenditure. Usually a lower level of capital expenditures was recommended by the regulator and to this level an incentive scheme that resembled a menu of contracts regulation model was applied. Quality of service was incentivised separately through performance targets. The targets for each DNO were individual and deviation from these resulted in company specific penalties and rewards that affected the total allowed revenue. This hybrid of incentive regulation implied a trade-off between Opex and Capex savings and the provision of non-tradable outputs such as service quality. Several studies proved an under-supply of service quality.

Given the trade-offs under RPI-X, an integrated benchmarking approach was suggested and implemented with RIIO[16]. RIIO is the new performance based model for setting the network companies' price controls in the UK and has been effective since January 2015. It builds on the previous RPI-X regime, but better meets the investment and innovation challenge. It is designed to promote smarter networks for a low carbon future that accommodate renewables, distributed generation, microgeneration and active demand (Ofgem, 2014). There are three separate price controls under RIIO: RIIO-T1 regulates the high voltage electricity and gas transmission for 2013-2021. RIIO-GD1 regulates gas distribution and RIIO-ED1 relates to the distribution, i.e. transport of electricity at lower voltage level to homes and commercial customers. RIIO-ED1 aims to provide strong incentives to ensure reliable supply. Moreover, Ofgem introduced a package of connection incentives to encourage the DNOs to provide a better service for connecting customers, in particular those connecting low carbon technologies and distributed generation. Under RIIO the DNOs are also incentivised to engage with local authorities, health providers, suppliers, other utilities and community groups in order to use the information they collectively hold on customers in vulnerable situations. RIIO is based on Totex, which means that the regulator adds all cost measures as well as some measure of monetary values of service quality (such as WTP) and network losses together to determine the allowed revenue. This mitigates the previously often critisised Opex/Capex bias.

However, even if RIIO is an improvement over previous schemes, not much has changed in terms of the methodology to estimate customer preferences and WTP for the diverse outputs. The incentive rates are based on estimates resulting from simple MNL models that assume homogeneous preferences or

[16] RIIO stands for Revenue=Incentives+Innovation+Outputs.

presume specific segments of consumer preferences. We show that the simplified assumptions employed by regulators such as Ofgem can have significant implications for the price controls.

Beauly - Denny Example

The Beauly-Denny High Voltage Transmission Line illustrates the evaluation challenges of investments in grid infrastructure. It is a high profile grid development project but not part of the price control, as it is not considered a national necessity for competition and consumer protection. However, in 2004 Ofgem created the Transmission Investment for Renewable Generation mechanism (TIRG) to fund transmission projects specific to the connection of renewables (Tobiasson, 2015) and granted the transmission companies the right to recover the project cost from their customers. The project involves the construction of a 220 kilometres long 400kV double circuit overhead transmission line set to replace the current single circuit 132kV transmission line (Tobiasson et al, 2015). The new line will mainly follow the same route as the old line, however changes in the use of land in the course of time required slight deviations from the old route. The landscape along the line is characterised by varying land uses including remote moorland, forests, river valleys, roads (A9), areas of famous tourist attractions (e.g. Stirling Castle and the Wallace monument) and some more populated areas. The project generated close to 20,000 objections from all over the world and led to the longest ever public inquiry in Scotland (Tobiasson et al, 2015). One of the reasons for this strong opposition was the demand for some sort of compensation. As Tobiasson et al. (2015) emphasise: 'the potential for financial compensation in transmission projects provides the basis for further research in the future. In principle, the redistribution of costs and benefits to reach a socially optimal outcome is a viable and seeming solution, yet its application in practice provides numerous obstacles. Further research is needed to establish the scope and perhaps the most efficient form of financial compensation, including the separation of assets.' North (1994) points out that the formal rules within the incentive regulation scheme have not been created to be socially efficient, but are rather designed to benefit those with bargaining power, i.e. the network companies with their capital and resource equipment and their experience. Even if Ofgem claims to focus on consumer interests, consumer participation is still low (Littlechild, ????). The investigation of customer preferences and their consideration in the decision process can be crucial for timely and effective decision making. When engaging customers it is important to keep in mind that communities are heterogeneous. They consist of many individuals and local firms with different preferences that might change over time (Tobiasson et al., 2015). Devine-Wright et al. (2010) for example find that communities that were more familiar with electricity transmission and generation facilities, such as renewable energy host communities, were generally more understanding and sympathetic to the idea of the new power-line.

Figure 5: Example Choice Card Experiment 3

Example Of Choice from 3rd Experiment (Domestic Non-LPN)

	As Now	Alternative 1	Alternative 2
Commitment to undergrounding overhead lines in areas of outstanding natural beauty and national parks for amenity reasons	1.5% of overhead lines per annum	1.5% of overhead lines per annum	5% of overhead lines per annum
Number of customers affected by major storms	180000 customers on average in a year	162000 customers on average in a year (10% better than now)	144000 customers on average in a year (20% better than now)
Number of major electricity sites across GB exposed to a potential flood risk	Around 1000 major electricity sites	Reduce to 900 major electricity sites	Reduce to 950 major electricity sites
Investment to reduce carbon emissions	Continue usage of current equipment and vehicles	Replace 5% per year with those using less polluting fuels	Replace 5% per year with those using less polluting fuels
Annual Electricity Bill	£200 (no change)	£205 (£5 increase)	£209 (£9 increase)
Choice (mark "X" in preferred option)	☐	☐	☐

In Experiment 3 customers faced six different choice cards and had to choose one of three alternatives (status quo, alternative A or B) each time. Customers have to trade-off changes in the resilience attributes against each other and against the change in the electricity bill.

Table H.1: Summary Statistics - Proportions

Social Class	
A	1%
B	16%
C1	29%
C2	23%
DE	30%
Age	
16-29	20%
30-49	41%
50+	38%
Children	
0	77%
1	15%
2	7%
3	1%
4	0%
5	0%
Annual income	
Under 10000	15%
10000-20000	23%
20001-30000	15%
30001-40000	11%
40001-50000	6%
50000-60000	2%
Over 60000	2%
Dont know	6%
Refused	20%

Table H.2: Number of Customers per DNO

dnoID	**Customers**
1	2,614,165
2	2.446,951
3	3,516,859
5	2,233,288
6	740,768
7	2,934,581
8	2,258,404
9	1,575,686
10	2,359,391
11	1,099,333
12	1,541,188
13	1,485,153
14	1,992,998
	26,798,765
Mean	2,061,443

Figure 6: Summary Statistics by DNO

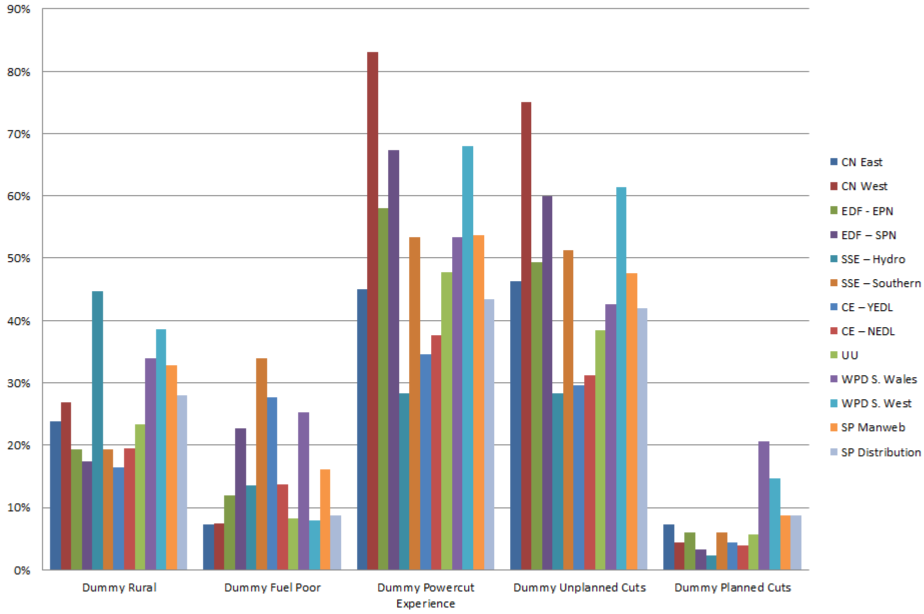


Figure 6 illustrates that there is heterogeneity in observables characteristics across DNOs. E.g.: while around 80 percent of the customers of CN West customers experienced power cuts, only 30 percent of the customers of SSE Hydro experienced power cuts within the last 12 months.

Table H.3: MNL with selected interactions

	MNL
Cost (£)	-0.109***
	(0.0115)
ASC2	-0.621***
	(0.0594)
ASC3	-0.763***
	(0.0593)
Less undergd (-1.5%)	0.254
	(0.376)
More undergd (+1.5%)	0.661***
	(0.199)
More storm damage (+10%)	-0.273***
	(0.0254)
Less storm damage (-10%)	0.170***
	(0.0240)
Less flood risk	0.0657***
	(0.0141)
Less CO2	0.382***
	(0.0446)
costXinc2	0.0303**
	(0.0114)
costXinc3	0.0381**
	(0.0132)
costXinc4	0.0462***
	(0.0130)
costXinc5	0.0613***
	(0.0132)
costXinc6	0.0700***
	(0.0167)
costXinc7	0.0750***
	(0.0146)
costXinc8	0.0201
	(0.0147)
costXinc9	0.0186
	(0.0142)
lCO2Xinc7	0.559***
	(0.141)
luXage3	-0.115
	(0.0634)
lCO2Xage2	-0.0173
	(0.0538)
lCO2Xage3	-0.143**
	(0.0539)
costXjob4	-0.0284***
	(0.00739)
costXjob5	-0.0200*
	(0.00976)
luXjob2	-0.435
	(0.381)
luXjob3	-0.515
	(0.377)
luXjob4	-0.490
	(0.378)
luXjob5	-0.366
	(0.376)
muXjob2	-0.454*
	(0.208)
muXjob3	-0.446*
	(0.203)
muXjob4	-0.495*
	(0.205)
muXjob5	-0.538**
	(0.204)
luXexpcuts	-0.258***
	(0.0639)
N	36036

Cluster robust standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The table lists the results from a MNL model with the selected interactions.

Table H.4: Posterior Summary Statistics by DNO: WTP Distributions as Resulting from M5

	1.5% More Undergrounding				10% Less Storm Damage			
	Mean	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
1	0.87	0.48	-0.88	2.19	0.83	0.49	-0.51	2.37
2	0.72	0.51	-0.39	3.01	0.78	0.40	-0.23	1.68
3	0.85	0.51	-0.60	2.51	0.87	0.48	-0.29	2.25
5	0.79	0.55	-0.70	2.71	0.80	0.46	-0.47	2.09
6	0.70	0.46	-0.69	2.25	0.78	0.41	-0.35	2.13
7	0.81	0.46	-0.25	2.25	0.85	0.49	-0.28	2.54
8	0.68	0.45	-0.63	1.94	0.78	0.48	-0.55	1.92
9	0.89	0.51	-0.17	2.33	0.82	0.46	-0.30	2.19
10	0.87	0.49	-0.71	2.73	0.94	0.48	-0.12	2.55
11	0.64	0.45	-1.00	1.55	0.71	0.42	-0.48	2.14
12	0.94	0.54	-0.92	2.70	0.89	0.49	-0.76	2.71
13	0.75	0.45	-0.47	2.34	0.77	0.43	-0.31	2.10
14	0.77	0.55	-1.15	2.40	0.81	0.41	-0.25	2.20

Table H.4 lists the summary statistic of the estimated individual posterior means by DNO. The individual means vary within and across DNOs.

Table H.5: Mean Customer Valuations of Improved Storm Resilience by DNO (£per annum per 1,000 customers affected).

DNO	Customers currently affected by storms		**WTP per 1,000 customers less affected by storms**			
	#	% of DNO	M1/M2	M3	M4	M5
1	180,000	7%	0.111	0.046	0.099	0.046
2	180,000	7%	0.111	0.046	0.097	0.044
3	260,000	7%	0.077	0.032	0.072	0.033
45	40,000	2%	0.498	0.206	0.434	0.201
6	80,000	11%	0.249	0.103	0.219	0.098
7	140,000	5%	0.142	0.059	0.128	0.061
8	60,000	3%	0.332	0.137	0.289	0.129
9	100,000	6%	0.199	0.082	0.175	0.082
10	60,000	3%	0.332	0.137	0.306	0.157
11	60,000	5%	0.332	0.137	0.285	0.118
12	140,000	9%	0.142	0.059	0.130	0.064
13	60,000	4%	0.332	0.137	0.294	0.128
14	60,000	3%	0.332	0.137	0.295	0.136

Table H.5 lists the WTP for 10 percent less customers affected by storms transformed into WTP per 1,000 customers. Based on table H.2, which lists the total number of customers per DNO, we calculated the share of customers that is currently affected by storms.

Table H.6: Aggregate WTP (£per annum per DNO) for Flood Resilience and CO2 Reductions

Less flood risk	M1/M2	M3	M4	M5
1	1,958,325	792,092	1,397,768	957,569
2	1,833,062	741,426	1,190,597	770,044
3	2,634,552	1,065,608	1,719,915	1,170,792
5	1,673,002	676,686	1,113,030	706,826
6	554,925	224,453	338,790	218,960
7	2,198,356	889,178	1,525,340	999,027
8L	1,691,817	684,296	1,062,593	615,674
9	1,180,379	477,433	767,176	524,252
10	1,767,469	714,895	1,242,095	784,420
11	823,533	333,098	467,021	254,174
12	1,154,536	466,980	805,908	529,988
13	1,112,559	450,001	727,922	491,089
14	1,492,996	603,878	967,346	560,622
Total	**20,075,511**	**8,120,026**	**13,325,502**	**8,583,439**
less CO2	M1/M2	M3	M4	M5
1	10,021,477	3,973,531	8,461,375	3,914,818
2	9,380,457	3,719,366	7,872,528	3,700,671
3	13,481,980	5,345,626	11,029,440	5,116,312
5	8,561,374	3,394,598	7,143,592	3,193,675
6	2,839,755	1,125,967	2,221,042	890,643
7	11,249,801	4,460,563	9,781,395	4,659,046
8	8,657,657	3,432,774	6,567,016	2,513,534
9	6,040,438	2,395,043	4,932,101	2,202,341
10	9,044,793	3,586,274	7,743,877	3,491,881
11	4,214,325	1,670,986	3,123,962	1,094,454
12	5,908,188	2,342,606	5,129,252	2,402,634
13	5,693,377	2,257,433	4,777,509	2,203,551
14	7,640,215	3,029,357	6,434,237	2,945,616
Total	**102,733,835**	**40,734,123**	**85,217,325**	**38,329,177**