I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16821

Alternative Models of Preference Heterogeneity for Elicited Choice Probabilities

Nathan Kettlewell
Matthew J. Walker
Hong Il Yoo

FEBRUARY 2024

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16821

Alternative Models of Preference Heterogeneity for Elicited Choice Probabilities

Nathan Kettlewell
University of Technology Sydney and IZA

Matthew J. Walker
Newcastle University

Hong Il Yoo
Loughborough University

FEBRUARY 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Alternative Models of Preference Heterogeneity for Elicited Choice Probabilities*

Discrete choice experiments (DCEs) often present concise choice scenarios that may appear incomplete to respondents. To allow respondents to express uncertainty arising from this incompleteness, DCEs may ask them to state probabilities with which they expect to make specific choices. The workhorse method for analyzing the elicited probabilities involves semi-parametric estimation of population average preferences. Despite flexible distributional assumptions, this method presents challenges in estimating unobserved preference heterogeneity, a key element in non-market valuation studies. We introduce a fractional response model based on a mixture of beta distributions. The model enables researchers to uncover preference heterogeneity under comparable parametric assumptions as adopted in conventional choice analysis, and can accommodate multiplicative forms of heterogeneity that make the semi-parametric method inconsistent. Using a DCE on alternative fuel vehicles, we illustrate the complementary roles of the parametric and semi-parametric approaches. We also undertake a separate analysis in which respondents are randomized to either a DCE employing a conventional choice elicitation format or a parallel DCE employing the probability elicitation format.

JEL Classification: C35, D12, D84, Q42, R41

Keywords: discrete choice experiment, probability elicitation, mixed logit, beta regression; willingness to pay

Corresponding author:
Nathan Kettlewell
University of Technology Sydney
Economics Discipline Group
15 Broadway
Ultimo NSW 2007
Australia

E-mail: Nathan.Kettlewell@uts.edu.au

* We would like to thank Denise Doiron for her instrumental role in designing the experiment and collecting the data used in this paper. She led our project team during her tenure as a Professor of Economics at UNSW Sydney, until her retirement. Two of us (Kettlewell and Yoo) had the privilege of being her PhD students and have witnessed her adherence to strict standards for authorship credit over the years. In keeping with these principles, she graciously declined authorship of this paper, citing her limited involvement in the manuscript's writing, which mostly occurred after her retirement. We extend our thanks to the UNSW Business School for providing financial support for data collection. Our study received ethics approval from the UNSW Human Research Advisory Panel (approval ID HC17239).

1 Introduction

The use of stated choice surveys—more commonly referred to as discrete choice experiments (DCE) in related literature—has become a staple in non-market valuation studies, overcoming initial skepticism regarding the reliability of stated preferences. Early adopters of DCEs in economics focused on consumer products featuring new and existing attributes (Beggs et al., 1981), following precedents in marketing science. The potential for comparing unobservable future scenarios with present conditions was subsequently recognized by researchers in environmental economics (Layton and Brown, 2000) and health economics (Ryan, 1999), leading to a rapid expansion of DCE studies in these areas over the past two decades. More recently, an increasing number of labor economists harness the experimental aspect of DCEs, which facilitates the measurement of workers' preferences in isolation from the influence of employers' preferences and market frictions (Wiswall and Zafar, 2018; Low, 2022; Koşar et al., 2022).

To reduce the cognitive burden on respondents, a choice scenario in DCEs typically comprises a succinct description based on a limited number of key attributes.[1] Manski (1999) emphasizes that the resulting scenario is likely to omit some attributes that respondents deem relevant to their decision-making, rendering it *incomplete* from their perspective. He argues that this incompleteness makes it more natural to ask respondents to state the probability with which they would make a particular choice, rather than to elicit their preferred choice as in traditional DCEs. Eliciting probabilities allows respondents' stated responses to reflect their expectations regarding omitted attributes.[2] Blass et al. (2010) demonstrate the empirical viability of this approach in a DCE study on the reliability of electricity supply, and advocate the use of the least absolute deviations (LAD) estimator which may be interpreted as a semi-parametric estimator of population average preferences. The implied semi-parametric model, however, does not nest several parametric models of interest to DCE studies, presenting a challenge to obtaining empirical results which are directly comparable between the two elicitation formats.

We introduce a flexible parametric approach for modeling interpersonal preference heterogeneity using data on elicited choice probabilities. Our approach takes the form of a fractional response model, which employs a beta distribution to account for stochastic measurement errors (*e.g.*, due to the rounding of small variations in probabilities) that may cause elicited probabilities to diverge from latent

[1]For example, 74% of DCE studies in health economics published in 2013-2017 use 6 or fewer attributes (Soekhai et al., 2019, Table 2).

[2]Juster (1966) was an early proponent of the expectations format, arguing that consumer purchase probabilities are inherently more informative than stated intentions to purchase.

choice expectations. The mean-variance parameterization of beta distributions, as proposed by Ferrari and Cribari-Neto (2004), allows us to specify a mean function that incorporates parametric models of preference heterogeneity which are widely used in the analysis of traditional choice data. These include models featuring a multiplicative form of heterogeneity (Train and Weeks, 2005; Fiebig et al., 2010) that the LAD estimator is unable to accommodate. As a result, for the first time, we can analyze elicited probabilities under assumptions about individual preferences that are directly comparable to those used in stated choice analyses. We leverage this aspect of our modeling approach in a study on consumer preferences for alternative fuel vehicles (AFVs), using a DCE dataset that has not previously been analyzed. In this DCE, participants were randomly assigned to the traditional choice format or the probability elicitation format, allowing us to evaluate the sensitivity of estimated preference structures to the response formats used.

Since the seminal empirical application by Blass et al. (2010), the LAD estimator of population average preferences has become a workhorse method in a small, yet growing number of studies adopting the probability elicitation format. Their areas of application illustrate the broad scope of contemporary DCE research and include environmental water quality (Herriges et al., 2011); land-use alternatives to mitigate biodiversity loss and climate change (Shoyama et al., 2013); voting for political candidates (Delavande and Manski, 2015); demand for electricity sources (Morita and Managi, 2015); long-term care insurance products (Boyer et al., 2017); workplace characteristics (Wiswall and Zafar, 2018); medical students' demand for general practice jobs (Pedersen et al., 2020); food choice (Scarpa et al., 2021); and migrants' location decisions (Koşar et al., 2022).[3]

The primary appeal of the LAD estimator lies in its ability to estimate population average preferences within the mixed logit framework (McFadden and Train, 2000) without imposing a specific distribution (*e.g.*, multivariate normal) on those preferences. However, this flexibility invites increased challenges in estimating the extent and source of interpersonal preference heterogeneity, which has been the focal point of mixed logit studies (Revelt and Train, 1998; Layton and Brown, 2000; Small et al., 2005). To facilitate further discussion, consider the distinction between *coefficient heterogeneity* and *scale heterogeneity* (Fiebig et al., 2010).[4] Infer-

[3]In addition to LAD, Scarpa et al. (2021) also apply a fractional logit estimator (Papke and Wooldridge, 1996). As we will discuss further in Section 2.3, the fractional logit estimator in this context cannot be interpreted as a maximum likelihood estimator (MLE) or a consistent quasi-MLE. By contrast, our approach is aligned with the MLE interpretation.

[4]Perhaps the most well-known type of mixed logit is one that accommodates heterogeneous preferences for observed attributes by specifying coefficients on those attributes as individual-specific random parameters. A parallel treatment of preferences for unobserved attributes requires the variance of the residual error term to be specified as a random parameter. The former captures

ring coefficient heterogeneity from the LAD results requires that the number of observations per respondent exceeds the number of heterogeneous preference parameters, a condition which is rarely satisfied in DCEs.[5] Further, since the LAD estimator assumes a linear log-odds specification, it is unable to accommodate scale heterogeneity, which is linked to interpersonal heteroskedasticity and introduces a multiplicative parameter to the implied log-odds. Yet scale heterogeneity is highly relevant for the original motivation to elicit choice probabilities: if individuals vary in their perceptions of scenario incompleteness, then the scale of the error term in the random utility function would be individual-specific. The importance of analyzing preference heterogeneity goes beyond purely technical considerations as the focus on average preferences might lead to the adoption of policies that are welfare-decreasing for large segments of the population. Additionally, willingness-to-pay (WTP) measures derived from the average preference parameters are generally different from the average of WTP distributions (Daly et al., 2012), and may not be as informative as the latter for cost-benefit analyses.

The fractional response model that we propose complements the LAD approach by allowing researchers to use elicited choice probabilities to uncover the population distributions of preference and WTP parameters. This is achieved under the same set of parametric assumptions concerning preference heterogeneity as those applied in the analysis of elicited choices. Our empirical analysis considers preference structures aligned with mixed logit models with normally distributed coefficients in the preference space (Revelt and Train, 1998) and the WTP space (Train and Weeks, 2005), as well as Generalized Multinomial Logit (GMNL) models that incorporate more flexible continuous mixture-of-normals distributions (Fiebig et al., 2010; Keane and Wasi, 2013). The well-known flexibility of beta distributions (Ferrari and Cribari-Neto, 2004) makes them an attractive option for representing reporting errors in elicited probabilities. When combined with the mixed logit representation of latent choice expectations, the resulting error distributions can exhibit varying degrees of dispersion, skewness, and modality between respondents, as well as across different choice scenarios within a respondent.

We apply the fractional response model to analyze preferences and WTP for AFVs in Australia. Purchasing a car involves a significant financial commitment, requiring the buyer to consider a complex array of individual-specific factors which cannot be fully incorporated into a generic choice scenario. Inherent consumer un-

coefficient heterogeneity, and the latter *scale heterogeneity*.

[5]This condition arises from the fact that (1) coefficient heterogeneity is identified by running an auxiliary OLS regression of LAD residuals on independent variables which are assumed to have heterogeneous coefficients; and (2) this auxiliary regression must be executed separately for each individual in the sample. We provide more discussion in Section 2.2.

certainty regarding future fuel technologies and potential changes in policy further adds to the complexity in decision-making. These factors make the DCE on AFVs an ideal domain in which to evaluate the convergent validity of probability elicitation and standard choice formats. We apply the same range of mixed logit specifications to the choice format data as are embedded in our fractional response model. This enables us to compare the two sets of results directly.

We find that, while the two response formats are generally qualitatively similar in their results, there are notable differences that often favor the choice probabilities format.[6] For example, we uncover a negative preference for liquid petroleum gas vehicles—which is consistent with market behavior in the time since our data were collected—in the choice probabilities data only. We also uncover richer patterns of preference heterogeneity regarding vehicle size and fuel type in the choice probabilities data; intuitively, preferences for such attributes are unlikely to be homogeneous considering interpersonal variations in factors such as lifestyle needs, openness to new technologies, and perceived social status. As is common in DCE studies, our limited number of observations per respondent precludes the use of the LAD results to identify coefficient heterogeneity. Nevertheless, our fractional response model estimates indicate a substantial amount of coefficient and scale heterogeneity, underscoring the importance of applying a method that can illuminate this aspect to complement the analysis of population averages. Finally, in our dataset we find that estimation is fragile with the stated choice data (but not choice probabilities data), necessitating computational compromises to achieve convergence. This hints at a practical advantage of eliciting choice probabilities: there is the potential to capture more nuanced variations in the response variable compared to simply observing how the preferred choice jumps discretely from one alternative to another as experimental stimuli (choice attributes) change.

Outside the realm of elicited choice probabilities in DCE settings, a distinct but related strand of literature elicits choice probabilities to investigate the impact of information interventions on belief updating (*e.g.*, Wiswall and Zafar, 2015; Ruder and Van Noy, 2017; Bleemer and Zafar, 2018; Miller et al., 2020). We are optimistic that our fractional response model will find useful applications in this context too, in particular for identifying how informational nudges might influence the distribution of beliefs in a population, not just the mean. There is also a large literature

[6]Similar to our DCE design, Herriges et al. (2011) and Shoyama et al. (2013) randomly assign participants to the choice format or the probability elicitation format. They also find that the estimated preference structures from the two formats exhibit qualitative similarities and quantitative differences. However, their analyses are based on the LAD estimator of a semi-parametric model for elicited probabilities and the MLE of a parametric model for elicited choices, making it difficult to ascertain whether the quantitative differences reflect differences in fundamental preference structures or modeling assumptions.

that relates expectations about states of nature to choice behavior in structural microeconomic models.[7] Of note for the present study is Hendren (2013), who elicits subjective probabilities about life events covered in insurance markets to infer and quantify the degree of private information in the market. Hendren approximates the distribution of subjective probabilities using a finite mixture of beta density functions, with the objective to account for heterogeneity in agents' beliefs. Although his analytical context fundamentally differs from ours, his likelihood function shares a related algebraic structure, because our estimation of the fractional response model is based on a continuous mixture of beta density functions.

The remainder of the paper proceeds as follows. In Section 2, we present the random utility framework as applied to traditional stated choice data and summarize how it can be generalized to incorporate different forms of preference heterogeneity. We then outline the linear model of subjective expectations due to Blass et al. (2010) and set out the proposed fractional response model of preference heterogeneity for analyzing elicited choice probabilities. In Section 3, we describe the data and experimental design on which we apply the different random utility models to estimate preferences for AFVs in Australia. In Section 4, we report the results of the empirical application. Section 5 concludes the paper.

2 Models

2.1 Mixed Logit Model of Discrete Choices

The mixed logit model, also known as the random coefficient logit model, is widely used in applied microeconomics to quantify unobserved preference heterogeneity. Studies such as McFadden and Train (2000) and Walker et al. (2007) show that the model can be employed as a tractable tool to approximate a variety of random utility maximization models, contributing to its wide application in discrete choice modeling across disciplines. We summarize major types of mixed logit models which are commonly employed in contemporary non-market valuation studies.

We address the typical setup of a discrete choice experiment (DCE) which has the same individual complete multiple choice tasks presenting different sets of alternative. Let $n \in \{1, 2, \cdots, N\}$ be the index of distinct individuals and $t \in \{1, 2, \cdots, T\}$ be that of decision tasks. Let $\boldsymbol{x}_{njt}$ and p_{njt} denote non-price and price

[7] Random utility models of the expectation-choice relation have been estimated in a variety of settings spanning personal financial outcomes (e.g., Dominitz and Manski, 1997; Hurd et al., 2004; McKenzie et al., 2013), health outcomes (e.g., Delavande, 2008; Tarozzi et al., 2014), returns to schooling (e.g., Arcidiacono et al., 2012; Zafar, 2013; Delavande and Zafar, 2019) and crime (e.g., Lochner, 2007). For an up-to-date survey at the time of writing, see Koşar and O'Dea (2023).

attributes of option $j \in \{1, 2 \cdots, J\}$ that individual n evaluates in task t.

Consider McFadden's random utility maximization model (McFadden, 1974). Individual n's preference for option j in task t is represented by a random utility function U_{njt}

$$U_{njt} = \boldsymbol{\beta}_n \cdot \boldsymbol{x}_{njt} - \lambda_n p_{njt} + \varepsilon_{njt}/\sigma_n \tag{1}$$

where $\boldsymbol{\beta}_n$ and λ_n represent utility weights on the observed attributes; ε_{njt} is an *i.i.d.* type 1 extreme value error term; and σ_n is a scale parameter that is inversely related to randomness in the individual's choice behavior. Conditional on the utility and scale parameters, the probability that the individual chooses option j in this task is given by

$$P_{njt} = \frac{\exp[V_{njt}]}{\sum_{i=1}^{J} \exp[V_{nit}]} \tag{2}$$

where $\exp[\cdot]$ indicates the exponential function; and the utility index $V_{njt} = \sigma_n \cdot (\boldsymbol{\beta}_n \cdot \boldsymbol{x}_{njt} - \lambda_n p_{njt})$ given equation (1). Of course, since any increase in σ_n is observationally equivalent to a proportionate shift in $\boldsymbol{\beta}_n$, the parameters must be normalized further to permit identification. As we will summarize shortly, the exact functional form of the index function V_{njt} varies across discrete choice models, depending on the approaches to normalization and the specification of interpersonal preference heterogeneity. The basic case of the multinomial logit model (MNL), for example, results from having the scale factor normalized to unity ($\sigma_n = 1$) and the utility weights specified as non-random coefficients that remain constant across all individuals ($\boldsymbol{\beta}_n = \bar{\boldsymbol{\beta}}$ and $\lambda_n = \bar{\lambda}$).[8]

The mixed logit model refers to a family of discrete choice models that accommodate interpersonal preference heterogeneity by specifying the attribute-specific utility weights, sometimes along with the scale factor, as random parameters (McFadden and Train, 2000). Perhaps the most well-known type of mixed logit model is the normal mixed logit model (Revelt and Train, 1998), which maintains the MNL setting of $\sigma_n = 1$ and specifies V_{njt} as

$$V_{njt} = \boldsymbol{\beta}_n \cdot \boldsymbol{x}_{njt} - \lambda_n p_{njt} \tag{3}$$

where $\boldsymbol{\beta}_n$ and λ_n are individual-specific draws from a multivariate normal distribution that represents the population distribution of preferences. Indeed, this type of model specification has been so widely used that the term "mixed logit model",

[8]One may still introduce interpersonal heterogeneity in the MNL framework by specifying the utility weights as a deterministic function of observed personal characteristics (Train, 2009, §3.3.1). In this case even the non-random utility weights need carry the index of individuals n. For the simplicity of notation we do not consider this case in the present section, and attach the index n only to those parameters that vary randomly between individuals.

without further qualification, is often understood to describe this particular specification rather than the family of models. For the clarity of the distinction between the family and the particular instance of it, we will refer to the latter and equation (3) as the NMXL model.

The focal point of non-market valuation is the willingness-to-pay (WTP) for non-price attributes, which is defined as $\boldsymbol{\omega}_n = \boldsymbol{\beta}_n/\lambda_n$. See, for example, Small et al. (2005). A key drawback of the NMXL model is that the implied WTP distribution does not have finite moments including the population average and variance, since the support of normal λ_n includes zero. Specifying λ_n as a positive-valued distribution, such as log-normal, is often considered unsatisfactory since the implied moments tend to take on implausible values.

The mixed logit model in the WTP space, proposed by Train and Weeks (2005), addresses these limitations by specifying the individual's preferences in terms of the WTP coefficients, rather than the utility weights, from the outset. Algebraically, their model can be seen as a special case of equation (1) that normalizes λ_n, instead of σ_n, to unity for all individuals. The resulting model is in the WTP space since this normalization leads to a utility index of

$$V_{njt} = \kappa_n(\boldsymbol{\omega}_n \cdot \boldsymbol{x}_{njt} - p_{njt}) \tag{4}$$

where $\boldsymbol{\omega}_n := \boldsymbol{\beta}_n/\lambda_n$ is the vector of WTP coefficients and $\kappa_n := \lambda_n \cdot \sigma_n$ is a composite parameter that captures heterogeneity in the taste for money, as well as in the scale of utility. The most popular approach is to complete the stochastic specification by modeling $\{\ln[\kappa_n], \boldsymbol{\omega}_n\}$ as individual-specific draws from a multivariate normal distribution, thereby linking the mean of $\boldsymbol{\omega}_n$ to the population average WTP. This distribution assumption represents a fundamental departure from the NMXL model, since one cannot derive normal $\boldsymbol{\omega}_n$ as the ratio of normal $\boldsymbol{\beta}_n$ and normal λ_n.[9] For ease of reference, we will refer to the mixed logit model in the WTP space in equation (4) as the NWTP model.

The GMNL-II model, developed by Fiebig et al. (2010), makes a fuller use of the notion that the mixed logit model can be normalized to accommodate scale heterogeneity alongside coefficient heterogeneity. The acronym is derived from the fact that it is a version of the more flexible, but also more exotic, Generalized Multinomial Logit (GMNL) model that we will summarize shortly. In GMNL-II, the scale factor σ_n is normalized by setting its population mean—rather than all of its individual-specific values as was the case with NMXL—to unity. The utility

[9] This follows from the general result that the ratio of two normally distributed random variables does not itself follow a normal distribution.

index in this instance is specified as

$$V_{njt} = \sigma_n(\boldsymbol{\beta}_n \cdot \boldsymbol{x}_{njt} - \lambda_n p_{njt}) \tag{5}$$

where $\boldsymbol{\beta}_n$ and λ_n are draws from a multivariate normal distribution as with NMXL, and $\ln[\sigma_n]$ from a normal distribution whose mean is calibrated to satisfy $\mathrm{E}[\sigma_n] = 1$.[10] As is the case under NMXL, the parameters $\boldsymbol{\beta}_n$ and λ_n in GMNL-II are interpreted as utility weights rather than WTP coefficients. GMNL-II nests NMXL in equation (3) as a special case with a degenerate σ_n, but it does not nest the NWTP model in equation (4) as a special case with a normalized price coefficient $\lambda_n := 1$. Unlike σ_n, the composite scale parameter κ_n in the NWTP model retains its mean as a free parameter.

Finally, the full version of GMNL generalizes equation (5) further by allowing scale heterogeneity to have distinct effects on the population mean of the utility weights and individual-specific deviations around the mean. To facilitate further discussion, let us decompose each individual's utility weights as $\boldsymbol{\beta}_n = \bar{\boldsymbol{\beta}} + \boldsymbol{\eta}_n$ and $\lambda_n = \bar{\lambda} + \nu_n$, where the overbar indicates the mean and $\boldsymbol{\eta}_n$ and ν_n the deviations. The utility index under GMNL is given by

$$V_{njt} = \sigma_n(\bar{\boldsymbol{\beta}} \cdot \boldsymbol{x}_{njt} - \bar{\lambda} p_{njt}) + [\gamma + \sigma_n(1-\gamma)](\boldsymbol{\eta}_n \cdot \boldsymbol{x}_{njt} - \nu_n p_{njt}) \tag{6}$$

where $\boldsymbol{\eta}_n$ and ν_n are draws from a zero-mean multivariate normal distribution; $\ln[\sigma_n]$ from a normal distribution whose mean is calibrated as with GMNL-II; and $\bar{\boldsymbol{\beta}}$, $\bar{\lambda}$, and γ represent non-random parameters to be estimated alongside these distributions. The only extra parameter compared to GMNL-II is γ, but this small addition lets the data speak for the extent to which interpersonal heterogeneity is better modeled as GMNL-II ($\gamma = 0$) or GMNL-I ($\gamma = 1$). In a "horse race" study which fits a variety of discrete choice models to 10 empirical data sets, Keane and Wasi (2013) conclude that GMNL either achieves the best fit or offers a close approximation to a better fitting model which is often computationally less tractable.

2.2 Linear Model of Elicited Log-Odds

In traditional DCEs, respondents are asked to select one of J different alternatives in each choice task. Manski (1999) and Blass et al. (2010) advocate an alternative response format that involves asking respondents to indicate the probability, in percentage terms, that they would make a particular choice. Given that most

[10] Suppose that $\ln[\sigma_n]$ follows a normal distribution with a mean of m and a standard deviation of τ. Then the mean of σ_n is equal to $\exp[m + 0.5\tau^2]$. As this formula implies, the mean normalization of σ_n in GMNL-II can be accomplished by treating τ as a free parameter and setting $m := -0.5\tau^2$.

DCEs limit the number of attributes describing each alternative to minimize information overload and respondent fatigue, the choice scenarios presented may be *incomplete* from the respondent's perspective. This incompleteness refers to the potential exclusion of certain relevant attributes in decision-making, requiring respondents to infer or imagine the missing information. Manski (1999) argues that, from a behavioral perspective, choice probabilities are a desirable response format that allows respondents to express their choice uncertainty due to those unspecified aspects. Blass et al. (2010) demonstrate that, from an empirical perspective, the use of choice probabilities is appealing as it facilitates semi-parametric estimation of mixed logit models. We now summarize their semi-parametric approach.

To facilitate further discussion, we assume $J = 2$ henceforth, as is the case with the empirical settings of Blass et al. (2010), as well as our own application below. The conditional choice probability in equation (2) implies that the log-odds of selecting alternative $j = 2$ is given by $\ln[P_{n2t}/P_{n1t}] = V_{n2t} - V_{n1t}$. Define $\Delta \boldsymbol{x}_{n2t} = \boldsymbol{x}_{n2t} - \boldsymbol{x}_{n1t}$ and $\Delta p_{n2t} = p_{n2t} - p_{n1t}$ as the differences in the observed attributes of the two alternatives under consideration.

With this new notation, we can express the log-odds under the NMXL model in equation (3) as

$$\begin{aligned} \ln[P_{n2t}/P_{n1t}] &= \boldsymbol{\beta}_n \cdot \Delta \boldsymbol{x}_{n2t} - \lambda_n \Delta p_{n2t} \\ &= (\bar{\boldsymbol{\beta}} \cdot \Delta \boldsymbol{x}_{n2t} - \bar{\lambda} \Delta p_{n2t}) + (\boldsymbol{\eta}_n \cdot \Delta \boldsymbol{x}_{n2t} - \nu_n \Delta p_{n2t}) \\ &= \bar{\boldsymbol{\beta}} \cdot \Delta \boldsymbol{x}_{n2t} - \bar{\lambda} \Delta p_{n2t} + \epsilon_{n2t} \end{aligned} \tag{7}$$

where the second equality applies the decomposition of multivariate normal $\boldsymbol{\beta}_n$ and λ_n into their population mean ($\bar{\boldsymbol{\beta}}$ and $\bar{\lambda}$) and individual-specific deviation components ($\boldsymbol{\eta}_n$ and ν_n), as introduced in the context of the GMNL-II model; and the third equality is derived by aggregating these deviations into a composite error term ϵ_{n2t}.

Suppose that individual n states that their probability of choosing alternative j in task t is equal to y_{njt}. The statistical analysis by Blass et al. (2010) is based on the core assumption that one may equate this elicited choice probability with the NMXL probability P_{njt}. Then one may estimate $\bar{\boldsymbol{\beta}}$ and $\bar{\lambda}$ in equation (7) conveniently by running a regression of $\ln[y_{n2t}/y_{n1t}]$ on $\Delta \boldsymbol{x}_{n2t}$ and Δp_{n2t}. Since the normal error components $\boldsymbol{\eta}_n$ and ν_n have a mean of zero, the composite error ϵ_{n2t} satisfies the zero conditional mean assumption and the ordinary least squares (OLS) estimator is consistent in this instance. Additionally, as the error components are symmetric, ϵ_{n2t} satisfies the zero conditional median assumption required for the consistency of the least absolute deviations (LAD) estimator. Unlike OLS, the LAD estimator offers robustness to outliers in the elicited log-odds, which may arise from the behavioral tendency to round off very small or large probabilities.

The zero conditional mean or median property of ϵ_{n2t} may be satisfied even if the underlying error components $\boldsymbol{\eta}_n$ and ν_n are not normal. This consideration adds a semi-parametric character to the OLS or LAD regression of elicited log-odds based on equation (7). The consistency of both estimators allows for any zero mean and symmetric distribution of $\boldsymbol{\eta}_n$ and ν_n, providing flexibility beyond the multivariate normality assumption of the NMXL model.

Despite the ease of estimation and semi-parametric advantage, the probability elicitation format and the log-odds regression approach are relatively underutilized in DCE studies. A notable drawback is that both OLS and LAD techniques are unsuitable in the presence of scale heterogeneity, which is characteristic of both the NWTP and the GMNL models. For illustration, consider the GMNL-II specification in equation (5), which implies following the log-odds

$$\begin{aligned}\ln(P_{n2t}/P_{n1t}) &= \sigma_n(\bar{\boldsymbol{\beta}} \cdot \Delta\boldsymbol{x}_{n2t} - \bar{\lambda}\Delta p_{n2t}) + \sigma_n(\boldsymbol{\eta}_n \cdot \Delta\boldsymbol{x}_{n2t} - \nu_n\Delta p_{n2t}) \\ &= \sigma_n(\bar{\boldsymbol{\beta}} \cdot \Delta\boldsymbol{x}_{n2t} - \bar{\lambda}\Delta p_{n2t}) + \sigma_n\epsilon_{n2t}\end{aligned} \tag{8}$$

where scale heterogeneity is represented by σ_n. Given that the individual-specific multiplicative factor is now attached to the mean coefficients, standard regression procedures can no longer be applied to estimate these coefficients. Furthermore the new composite error term $\sigma_n\epsilon_{n2t}$, as a product of two random variables, may not have the conditional mean or median of zero, even if ϵ_{n2t} does when considered in isolation. The inability to accommodate scale heterogeneity warrants a concern in light of empirical findings suggesting that this aspect accounts for a substantial share of interpersonal heterogeneity in discrete choice data.[11] Moreover, scale heterogeneity is highly relevant given the notion of incomplete scenarios that motivates the elicitation of choice probabilities. Different individuals may perceive an identical scenario as incomplete to varying extents, and these perceptual differences imply that the relative importance of the error term in the random utility function varies from individual to individual, hence scale heterogeneity.

Beyond its apparent limitations in addressing scale heterogeneity, the log-odds regression model in equation (7) falls short in providing several key parameter

[11] In an analysis of model fit across 10 empirical datasets, Fiebig et al. (2010) find that models which account for scale heterogeneity (such as GMNL or its variant that preserves scale heterogeneity while assuming coefficient *homogeneity*) outperform NMXL in every case. This finding aligns with an earlier study by Hensher et al. (1998) in the basic MNL framework, which concluded that discrepancies between stated and revealed preference data could be more parsimoniously modeled by introducing a data-specific scale factor, instead of allowing the entire coefficient vector to shift between the two data sources.

estimates crucial in DCE studies. As noted earlier, these studies are usually motivated by non-market valuation questions, which relate to WTP measures $\boldsymbol{\beta}_n/\lambda_n$. One can use the log-odds regression coefficients to evaluate the *average* decision maker's WTP as long as the *average* is defined in terms of utility parameters: the WTP takes the form of $\bar{\boldsymbol{\beta}}/\bar{\lambda}$ in this case. But the results are difficult to interpret in relation to the population distribution of WTP because this ratio does not reflect the average or median of the WTP distribution *per se*, which is more pertinent to non-market valuation: the mean or median ratio of two random variables does not equate to the ratio of their respective means or medians.

Finally, empirical research often aims to quantify not just the central tendency of preference parameters, but also their population distributions (Layton and Brown, 2000; Small et al., 2005). Employing log-odds regression to study population distributions requires that the number of choice observations per individual, denoted as T in our notation, exceed the combined dimension of $\boldsymbol{x}_{njt}$ and p_{njt}.[12] Many DCE studies, including Hackbarth and Madlener (2013, 2016) from which we borrow our design elements (see Section 3), do not meet this requirement: The common practice of adopting a small T to minimize respondent fatigue often makes it infeasible to apply the log-odds approach to measure population distributions.[13]

2.3 Fractional Response Model of Elicited Choice Probabilities

To complement the semi-parametric estimator using the elicited log-odds, it would be advantageous to employ a parametric model of individual heterogeneity for the choice probabilities data. Ideally, such a model should incorporate comparable assumptions regarding individual preference heterogeneity as the widely-used parametric mixed logit models in the DCE literature.

Our objective is to formulate this type of model while accounting for evaluative or reporting errors that may introduce discrepancies between a respondent's latent and stated choice probabilities. The integration of these errors into the assumed decision making process is a cornerstone of McFadden's random utility model, especially in the context of its application to traditional DCE analysis. However, the existing literature does not provide a framework that researchers can readily

[12]To estimate individual-specific $\boldsymbol{\eta}_n$ and ν_n, one should run an auxiliary regression of residuals (estimates of ϵ_{n2t}) on $\Delta\boldsymbol{x}_{n2t}$ and Δp_{n2t}, separately for each individual n. The dimensionality condition relative to T arises from the standard rank condition required for independent variables to identify regression coefficients.

[13]Note that $\boldsymbol{x}_{njt}$ often encompasses more variables than a simple count of distinct product attributes. For example, a qualitative attribute with K levels (*e.g.*, vehicle size options like small, medium, or large) may require representation through $K-1$ distinct dummies rather than just one variable. In labeled choice experiments, an additional alternative-specific constant is needed for each labeled alternative, further expanding the dimension of $\boldsymbol{x}_{njt}$.

employ to address this modeling issue when engaging in the analysis of elicited choice probabilities. We aim to address this gap.

As with the preceding subsection, let P_{n2t} denote the respondent's latent choice probability for alternative 2, which incorporate their subjective evaluation of the incomplete aspects of a choice scenario. Differently from the log-odds regression approach, however, we explicitly consider the elicited choice probability y_{n2t} as a noisy manifestation of this latent preference. This approach aligns with the familiar decomposition of a latent dependent variable into systematic evaluation and response error components in behavioral choice modeling. A beta distribution with a mean of P_{n2t} offers a particularly attractive stochastic specification of y_{n2t} to account for the random evaluative noise. This distribution has support on the unit interval, as required for representing elicited probabilities. Moreover, it is well-known for its ability to take on a variety of shapes depending on the data, including bimodality and left or right skewness, without requiring the researcher to impose such features as *a priori* assumptions.

For concreteness, let $\boldsymbol{\theta}_n$ denote the parameters that represent individual preferences in a particular model, *i.e.*, those terms on the right-hand of V_{njt} in equations (3) to (6) other than the observed attributes. We assume that, conditional on $\boldsymbol{\theta}_n$, y_{n2t} follows a beta distribution with a mean of P_{n2t} and a variance of $P_{n1t}P_{n2t}/(1+\phi)$, where $P_{n1t} = 1 - P_{n2t}$ and ϕ is the distribution's precision parameter to be estimated. The conditional density of y_{n2t} can be then specified as

$$l_{nt}[y_{n2t}; \boldsymbol{\theta}_n, \phi] = \frac{\Gamma[A_{nt} + B_{nt}]}{\Gamma[A_{nt}]\Gamma[B_{nt}]} y_{n2t}^{A_{nt}-1} y_{n1t}^{B_{nt}-1} \tag{9}$$

where $\Gamma[\cdot]$ denotes the gamma function, $y_{n1t} = 1 - y_{n2t}$, $A_{nt} = \phi P_{n2t}$, and $B_{nt} = \phi P_{n1t}$. This conditional density function is equal to the density function of the usual beta regression model, except that θ_n includes random, as opposed to fixed, parameters.[14] While it might seem natural to have the precision parameter ϕ specified as an individual-specific random parameter ϕ_n akin to $\boldsymbol{\theta}_n$, our experience in the empirical analysis below indicates that such a model specification is difficult to identify empirically. Considering that the variance of y_{n2t} under the beta distribution is $P_{n1t}P_{n2t}/(1+\phi)$, where each P_{njt} is a function of $\boldsymbol{\theta}_n$, interpersonal preference heterogeneity inherited from mixed logit already induces this variance to exhibit interpersonal heteroskedasticity. We speculate that this makes it difficult to disentangle the distinct role of possible heterogeneity in ϕ, and specify it as a non-random parameter.

The utility index V_{njt} which underpins P_{n2t} and P_{n1t} in equation (9) can incor-

[14]Put another way, just as mixed logit models can be seen as random coefficient MNL models, this fractional response model can be seen as random coefficient beta regression models.

porate any mixed logit structure, including NMXL, NWTP, GMNL-II, and GMNL specifications reviewed above. This enables the use of elicited probabilities in the estimation of the same set of preference parameters as one could do with discrete choice data. In our empirical analysis below, we exploit this feature to draw fuller comparisons of preference structures estimated from a probability elicitation experiment to those from a traditional DCE, where both types of experiments were based on the same set of choice scenarios. As detailed in Appendix A, the unconditional likelihood function for a fractional response model based on equation (9) has the same algebraic form as that for the corresponding mixed logit model, except that the beta conditional density takes the place of the logit conditional density in each choice scenario. Similarly as with the case of mixed logit (Train, 2009, S6.7), we apply the method of maximum simulated likelihood (MSL) to maximize a simulated analogue to this likelihood function with respect to parameters characterizing the population distribution of $\boldsymbol{\theta}_n$ and the precision parameter ϕ.

Our modeling approach is closely related to two preceding studies. Scarpa et al. (2021) pursue a similar goal of estimating the mixed logit type of individual heterogeneity from elicited choice probabilities, employing a fractional logit model inspired by Papke and Wooldridge (1996). In our notation, they use the fractional logit density of the form $P_{n1t}^{y_{n1t}} P_{n2t}^{y_{n2t}}$ in the place of our equation (9). This approach, however, falls short of a formal statistical justification: The model specification is incomplete from the perspective of MSL as the fractional logit density does not fully specify a data generating process for y_{n2t}, and does not result in a consistent quasi-maximum likelihood estimator either as it falls outside Papke and Wooldridge's intended scope for the fractional logit method.[15] In an analysis of subjective beliefs over life events, Hendren (2013, pp.1750-1751) employs a *finite* mixture of beta density functions to model the belief distributions, leveraging the flexibility of beta distributions. His modeling framework is markedly different from ours since he directly considers probabilities as primitives, whereas our primitives are preference parameters which, together with observed attributes, generate probabilities. Nevertheless, our modeling approach is broadly aligned with his in the sense that integrating out $\boldsymbol{\theta}_n$ from equation (9) to obtain the unconditional likelihood function results in a *continuous* mixture of beta density functions.

[15]Papke and Wooldridge (1996) focus squarely on estimating a regression function for fractional responses which does not involve any random parameter. In this setting the fractional logit density leads to a quasi-maximum likelihood estimator of non-random regression coefficients which is consistent for a variety of true regression functions. This result is, however, irrelevant to the estimation of the population distribution of random parameters $\boldsymbol{\theta}_n$.

2.4 Flexible Models of Choices and Expectations

Perhaps the most well-known semi-parametric estimator of a random utility discrete choice model is the maximum score (MS) estimator due to Manski (1975, 1985). In terms of our notation for the log-odds approach in equation (7), this method provides a consistent estimator of the ratio $\boldsymbol{\delta} = \bar{\boldsymbol{\beta}}/\bar{\lambda}$ under a relatively mild assumption that $(P_{n2t} - P_{n1t}) > 0$ if and only if $(\boldsymbol{\delta} \cdot \Delta\boldsymbol{x}_{n2t} - \Delta p_{n2t}) > 0$. As a semi-parametric estimator, the MS method offers greater flexibility than the OLS and LAD methods for log-odds regression due to its capability to accommodate non-logit functional forms of P_{njt}. Unlike OLS and LAD, the MS method provides robustness against any form of scale heterogeneity attributed to $\sigma_n \in (0, \infty)$, because the inequality $(\boldsymbol{\delta} \cdot \Delta\boldsymbol{x}_{n2t} - \Delta p_{n2t}) > 0$ is equivalent to $\sigma_n(\boldsymbol{\delta} \cdot \Delta\boldsymbol{x}_{n2t} - \Delta p_{n2t}) > 0$. As Yan and Yoo (2019, S2.3) discuss in detail, the latter feature implies that the MS estimator is consistent for the NWTP and GMNL-II models as well as the NMXL model, along with their variants where the coefficients follow symmetric but non-normal distributions.[16] A well-known limitation of the MS estimator, however, is that it lacks point identification unless the variable linked to the normalized coefficient (in this case, Δp_{n2t}) exhibits continuous and unbounded support.[17] In all DCE designs, each attribute consists of a small finite number of distinct levels, implying that the MS estimator can only achieve set identification in the resulting data.

The original version of the MS estimator is designed for binary choice data and can be implemented by finding the values of $\boldsymbol{\delta}$ that maximize Kendall's rank correlation coefficient between $d_{n2t} - d_{n1t}$ and $(\boldsymbol{\delta} \cdot \Delta\boldsymbol{x}_{n2t} - \Delta p_{n2t})$, where d_{njt} is a choice indicator equal to 1 if respondent n chooses alternative j in task t and 0 otherwise. Following a conceptual development by Manski (1999), Blass et al. (2010) adapted this method for the expectations context by using the difference between elicited choice probabilities, denoted as $y_{n2t} - y_{n1t}$, in place of the binary choice indicators. To complement our parametric analysis, in which we estimate the same parametric models of preference heterogeneity for both standard DCE and choice probabilities data, we apply the respective procedures to compute MS estimates for the two data types. Thus, we can draw parallel semi-parametric comparisons.

[16] However, the MS estimator may not offer similar robustness to the full GMNL in equation (6), where the scale factor influences the mean coefficients and individual-specific deviation terms by different proportions due to γ. When expressed in terms of unobservables rather than choice probabilities, the consistency condition for the MS estimator is a zero conditional median assumption about a composite error term that incorporates unobserved preference heterogeneity and idiosyncratic disturbance. Unlike for GMNL-II, where the scale factor σ_n does not affect the question of whether $V_{n2t} - V_{n1t}$ is greater than 0, it does impact this directional comparison under GMNL. The implied unobserved preference heterogeneity is captured by $\sigma_n(1-\gamma)(\boldsymbol{\eta}_n \cdot \Delta\boldsymbol{x}_{n2t} - \nu_n \Delta p_{n2t})$, where the multiplication by σ_n means that it may not exhibit zero-mean symmetry, even if $\boldsymbol{\eta}_n$ and ν_n do.

[17] While one may choose any parameter in $\bar{\boldsymbol{\beta}}$ for normalization, $\bar{\lambda}$ is a convenient choice as it allows interpretation of the identified coefficients as reflecting the average decision maker's WTP.

3 Data and Experimental Setup

3.1 Overview

We compare alternative models of preference heterogeneity using data from a DCE that has not previously been analyzed. Our primary interest is in modeling elicited choice probabilities to compare different fractional response model specifications and the LAD approach. Nevertheless, in our DCE, respondents were randomly assigned to complete either probability elicitation tasks or standard choice tasks. This feature enables us to extend our analysis to evaluate the sensitivity of estimated preferences to the different response formats.

Our DCE concerns consumer preferences for alternative fuel vehicles (AFVs) in Australia. This is a market where choice uncertainty is potentially important since the market is emergent, new cars are an irregular purchase, and there are likely to be many factors outside our choice experiment that will influence preferences and increase the liklihood of choice uncertainty. Our data were collected in 2017, at which point AFV penetration was limited but growing in Australia. For example, less than 0.2% of new cars sold were electric or plug-in hybrid in 2017 (Electric Vehicle Council, 2023) but in the 12 months to Q3 2023 they accounted for 8.5% of sales, with a further 10.3% of sales being traditional hybrids (Australian Automobile Association, 2023).

3.2 Choice Set Design and Attributes

Following similar studies in Germany (Hackbarth and Madlener, 2013, 2016), we consider the following attributes on consumer choices: vehicle size; fuel type; fuel cost; CO2 emissions; recharge time; refill availability; driving range (maximum before re-fuel); and purchase price. Vehicle size was not included in the German studies, but this is an important feature in the Australian market, and is included in related Australian studies (Abdoolakhan, 2010; Beck et al., 2013). The levels and conditions for each attribute are specified in Table 1.

Our choice sets were constructed using a D-efficient fractional factorial block design. The reference model for the efficiency measure was multinomial logit with a coefficient vector of zeros. We generated 64 binary choice sets segmented into eight blocks. Each participant therefore saw 8 different scenarios in random order and each scenario presented a choice between two alternatives without an opt-out option. In addition to the conditions listed in Table 1, we also ensured that the fuel type varied in each scenario.

To compare the two response formats, we used a between-person design. Each

Table 1: DCE attributes and levels

Attribute	Levels	Conditions
Price	In $000's: 20, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80	See note below
Size	Small, Medium, Large	None
Fuel type	Conventional (CV), Hybrid electric (HEV), Liquid petroleum gas (LPG), Hydrogen fuel cell (FCEV), Battery electric (BEV), Plug-in hybrid electric (PHEV)	Always vary in choice sets
Fuel cost	$6.50, $13, $19.50	None
CO2 Emissions	0%, 50%, 75%, 100% of average car	If CV, LPG or HEV then 50%, 75% or 100%; If FCEV, BEV or PHEV then 0%, 50% or 100%
Recharge time	NA, 10 min, 1 hr, 6 hrs	If CV, LPG, FCEV or HEV then NA; If BEV or PHEV then 10 min, 1 hr or 6 hrs
Refill availability	20%, 60% and 100% of all stations	If CV, HEV or LPG then 60% or 100%; If FCEV, BEV or PHEV then 20%, 60% or 100%
Driving range	100km, 400km, 700km, 1000km	If CV, HEV, LPG, FCEV or PHEV then 400km, 700km or 1000km; If BEV then 100km, 400km or 700km

Notes: Relevant price levels vary by fuel type and vehicle size as listed in the following parentheses. CV and LPG: Small (20, 25, 30); Medium (30, 35, 40); and Large (30, 45, 60). HEV and PHEV: Small (30, 35, 40); Medium (40, 45, 50); and Large (40, 55, 70). BEV and FCEV: Small (40, 45, 50); Medium (40, 45, 50); and Large (40, 65, 80).

participant had a 5/7 chance of being allocated to the standard DCE asking them to state their preferred choice in each scenario, and a 2/7 chance of being allocated to the alternative format asking them to state their choice probabilities instead.[18] The instructions given to participants in each treatment, and an example choice set, are provided in Appendix B.

For DCEs to be informative, it is important that the choice set includes at-

[18]Our allocation rule gave less weight to the probability elicitation format because, at the time of data collection, we did not anticipate being able to use elicited choice probabilities to estimate preference heterogeneity. Since LAD is limited to estimating population average preferences given our design, we chose not to assign equal weights to both elicitation formats. Our subsequent development of the fractional response model made the estimation of preference heterogeneity feasible.

tributes and levels that are real-world relevant (*i.e.*, are calibrated to actual or potential consumer choices). We chose the set of AFV types based on Hackbarth and Madlener (2013), but decided against including liquid natural gas and 20% biofuel vehicles as these markets were practically non-existent, and seemed unlikely to develop, in Australia as of 2017 (this has indeed been the case in the years since our experiment). We added liquid petroleum gas (LPG) as this was a popular fuel alternative in 2017, although since then its popularity has waned (Laverick, 2023). Our attribute levels for CO2 emissions, recharge time, driving range and recharging time are in line with Hackbarth and Madlener (2013). For fuel cost, we used public data from the Australian Bureau of Statistics on fuel usage and from the Australian Petroleum Institute on fuel prices to determine a typical fuel cost of around AUD$13/100km for passenger vehicles. We used this as a middle-price and $\pm 50\%$ for low- and high-price. Finally, our levels for vehicle costs (in $AUD) were mainly informed by Abdoolakhan (2010, Fig. 4.2), which were calibrated using a focus group and a pilot study. We started with a range of prices for new, automatic transmission CV, LPG and HEV vehicles, adjusted the prices to 2017 dollars, and added prices for other categories based on an informal search of popular models. We then checked our price ranges with Hackbarth and Madlener (2013) and Beck et al. (2013) and confirmed the ranges were broadly similar.

Our scenarios were tested on a pilot sample of 49 people recruited from the same marketing panel used for our main dataset. Our pilot data supported our design choices, with a large degree of variation in options selected across the scenarios, and sensible coefficient estimates.

3.3 Sample Recruitment and Characteristics

Participants in our study were drawn from Qualtrics marketing research panels, and completed the allocated type of decision task and other survey questions online. Our recruitment process targeted people aged 25-70 years and a balanced pool of men and women. Our final sample is 60% female and has a mean age of 45.9 years, with 6 people outside the target age group. To ensure participants were engaged, we included an attention filter and people who failed this screening question are not included in our estimation sample.[19]

Table 2 provides demographic information on our sample, as well as participants' vehicle purchasing history and intentions, and attitudes towards climate change. Column 1 includes population benchmarks derived from population weighted

[19]The attention filter was a four item matrix asking how satisfied (5-point scale) they are with their life, finances, and neighbourhood, and then the attention filter instructing them to select 'dissatisfied'. 127 people (14.4%) failed the attention filter and were omitted.

Table 2: Sample descriptive statistics

	Population (HILDA)	All respondents	Choice respondents	Probability respondents
Female	0.511	0.604	0.592	0.634
Age	45.9	48.3	48.7	47.7
Australian born	0.683	0.709	0.703	0.723
Lives major city	0.590	0.674	0.665	0.703
Married/de-facto	0.711	0.613	0.638	0.624
Household size	3.05	2.568	2.569	2.555
Num. dependent children	0.771	0.602	0.593	0.625
Graduate	0.338	0.332	0.358	0.302
Employed	0.715	0.555	0.552	0.564
Retired	0.139	0.230	0.236	0.213
Household income:				
$0-≤$50k	0.139	0.321	0.328	0.307
$50k-≤$100k	0.267	0.514	0.512	0.520
$100k+	0.594	0.164	0.160	0.173
Owns an AFV	–	0.033	0.029	0.043
Car last purchased:				
0-5 years	–	0.744	0.726	0.787
5-10 years	–	0.146	0.158	0.134
10+ years	–	0.082	0.093	0.064
Car intended next purchase:				
Within 3 years	–	0.440	0.433	0.461
3-6 years	–	0.188	0.195	0.168
6+ years	–	0.077	0.081	0.068
Unsure/never	–	0.295	0.291	0.304
Agree about climate change that:				
It matters	–	0.679	0.716	0.668
Vehicles a main cause	–	0.751	0.768	0.792
Number of respondents	–	755	525	202

Notes: Population statistics are derived from an external data source (HILDA). All respondents refer to all participants in our DCE. Choice (probability) respondents refer to those who were randomly allocated to complete standard choice (probability elicitation) tasks. Every respondent evaluated 8 scenarios presenting two alternative fuel vehicles at a time.

data from the Household Income and Labour Dynamics in Australia Survey (HILDA) (Summerfield et al., 2022).[20] Although our sample is not representative of Australians aged 25-70 in 2017, it is broadly similar across many dimensions, including age, heritage, where they live, couple status, household composition, and education. The employment rate is lower in our sample than the population (55% versus

[20]The HILDA Project was initiated and is funded by the Australian Government Department of Social Services (DSS) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this paper, however, are those of the author and should not be attributed to either DSS or the Melbourne Institute.

72%) and our sample has lower household income, although this could be partly due to income being measured more comprehensively in the HILDA survey.

We also observe a high degree of engagement with the auto market: 74% of participants had purchased a car in the last five years and 44% intended to purchase a car in the next three years. The rate of AFV ownership[21] is low (3%), but in line with the population rate presented by another source (Australian Bureau of Statistics, 2018). Lastly, we observe that most people (68%) agree that climate change matters and that vehicles are a main cause (75%) which suggests they should value the environmental benefits of AFVs.

We note that our two treatment samples (*i.e.*, those who completed standard choice tasks and those who completed probability elicitation tasks) are highly comparable in terms of each characteristic. None of the differences are large in magnitude or statistically significant (p-values > 0.1 on all differences), demonstrating that our randomization worked.

4 Results

4.1 Elicited Choice Probabilities

We first present the raw distribution of elicited choice probabilities across 1,616 data points for 202 participants.[22] Each data point making up Figure 1 represents a subjective probability of choosing the second alternative in a given choice scenario. In about 80% of cases, we observe interior probabilities rather than 0 or 100%, which suggests that choice uncertainty is prevalent in this setting. The remaining fifth are almost equally split between the two boundary points, as one may expect given the random assignment of choice attributes to alternatives. As usual with stated probabilities (see Manski, 2004; Manski and Molinari, 2010; Blass et al., 2010), we observe signs of rounding behavior which induces spikes at multiples of 5%. For model estimation, we follow Blass et al. (2010) and recode the boundary responses of 0 and 100% as 0.1 and 99.9%: Similar to their log-odds regression, our fractional response model requires all data points to be interior probabilities.[23] Together with potential rounding behavior, this recoding makes it ap-

[21] Our definition here includes duel fuel (LPG/petrol) vehicles

[22] We drop 8 observations where stated probabilities do not add up to 100 within choice sets.

[23] Although rounding near the boundaries of the [0, 100] interval is a problem for OLS estimation of the linear model, assuming symmetric preferences circumvents the inference problem and permits estimation of the linear median regression model by LAD. The fractional response model addresses rounding directly through the scale parameter of the beta regression function, ϕ, which accounts for dispersion in how precisely respondents state their true probability in the survey. An alternative approach to analyze variables that represent probabilities with an excess proportion of

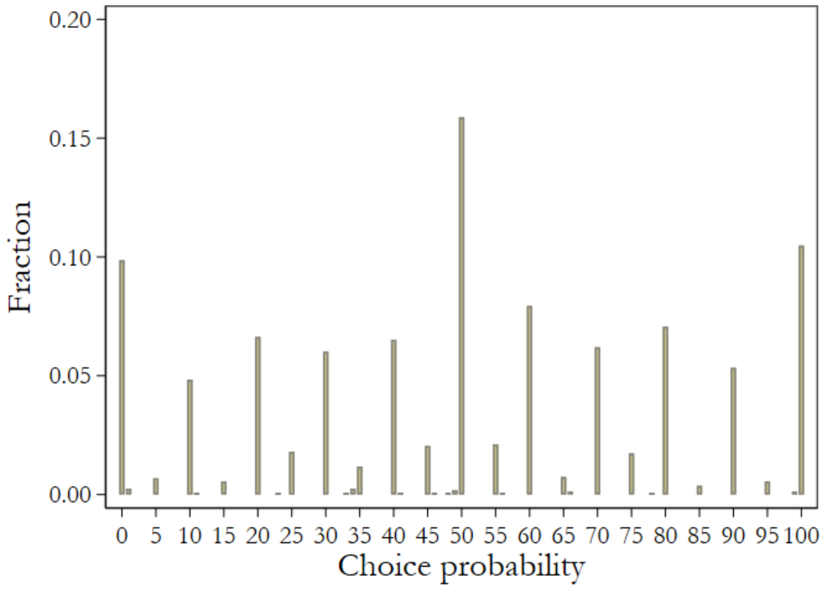

Figure 1: Histogram of choice probabilities for the second vehicle in each scenario.

propriate to view each data point as a combination of a true latent probability and measurement error. The LAD estimator of Blass et al. (2010) addresses this issue implicitly by virtue of its usual consistency condition that allows for measurement error that does not affect the location of the conditional median. Our fractional response model addresses it more explicitly by specifying each elicited probability as a draw from a beta distribution, the mean of which is the latent probability.

4.2 Population Average Preferences

We now use elicited choice probabilities in Figure 1 to study preferences for AFVs in Australia. To benchmark the main method used in current literature, we apply the LAD estimator of equation (7). We are interested in how these LAD results compare to fractional response models based on equation (9), which allow us to incorporate alternative parametric forms of preference heterogeneity presented in Section 2.1. We compute the MSL estimates of these fractional models by using 500 shuffled Halton draws to simulate sample likelihood functions.

In our DCE, each respondent evaluates 8 different scenarios, where each scenario is described by eight attributes in Table 1. We specify all six numeric attributes (price, fuel cost, CO2 emissions, recharge time, refill availability and driving range) as continuous variables, and expand the two qualitative attributes (ve-

zeroes and ones is the zero-one inflated beta regression model. However, this model discards the observations with zeroes and ones and so truncates the sample based on the observed responses.

hicle size and fuel type) into dummy variables. Since there are three vehicle sizes and six fuel types, even this parsimonious specification entails a minimum of 13 preference coefficients (6 + 2 + 5) to estimate.[24] We thus face a typical data environment for DCE studies, where the number of preference coefficients exceeds that of observations per respondent. This precludes the use of the LAD estimates to further identify the extent of interpersonal heterogeneity in each coefficient.[25] While we estimate this heterogeneity as part of our fractional response models, for now we focus on population average coefficients—that is, $\bar{\boldsymbol{\beta}}$ and $\bar{\lambda}$ as per our notation in Section 2—which can be estimated by both the benchmark method and our approach. We will discuss results that incorporate interpersonal heterogeneity in the next two subsections.

Table 3 presents alternative estimates of population average preference coefficients $\bar{\boldsymbol{\beta}}$ and $\bar{\lambda}$. The first column reports the benchmark LAD estimates based on the log-odds specification in equation (7). Subsequent columns present fractional response models that incorporate successively richer preference structures. The second column reports the fractional response model of multinomial logit preferences (F-MNL). This baseline model incorporates the standard MNL model—which assumes the same preference coefficients, *i.e.*, $\boldsymbol{\beta}_n = \bar{\boldsymbol{\beta}}$ and $\lambda_n = \bar{\lambda}$ for all individuals n—as the expected value P_{n2t} of the beta distribution in equation (9). The third column (F-NMXL) extends this specification to incorporate interpersonal preference heterogeneity as assumed by the NMXL model in equation (3). The fourth (F-GMNL-II) and fifth (F-GMNL) columns generalize the NMXL structure further to accommodate interpersonal scale heterogeneity, following GMNL-II and GMNL specifications in equations (5) and (6).

We obtain mostly the same results in terms of signs and significance across all model specifications. As one may expect from standard consumer theory and anecdotal evidence, in all models the average consumer dislikes a higher price or fuel cost, and has a preference for a greater refill availability and driving range (p-values < 0.01). We find suggestive evidence of pro-environmental preferences in our sample, as the average consumer's utility decreases significantly with greater CO2 emissions (at the 5% level in the LAD model and at the 1% level in the fractional response models).[26] Recharge time is also a source of disutility, but it is only

[24]We exclude dummies for "small" vehicle size and "conventional" fuel type which act as the reference categories.

[25]One solution to this underidentification problem is to assume that only a subset of preference coefficients exhibit heterogeneity. However, this solution is difficult to justify in light of common findings from DCE studies, which report substantial heterogeneity in tastes for most, if not all, included attributes. As reported in Appendix C (Table C1), the results from our fractional response models align with these common findings.

[26]The evidence is only suggestive as the enactment of environmental policies, such as emissions

Table 3: Population-level preference estimates for AFVs using choice probabilities.

	LAD	F-MNL	F-NMXL	F-GMNL-II	F-GMNL
Medium size	0.067	0.076	0.069	0.150**	0.080
	(0.071)	(0.052)	(0.069)	(0.060)	(0.062)
Large size	0.080	0.123**	0.110	0.054	0.035
	(0.073)	(0.056)	(0.074)	(0.086)	(0.070)
Liquid Petroleum Gas	-0.129	-0.080	-0.098	-0.472***	-0.309***
	(0.109)	(0.073)	(0.090)	(0.087)	(0.098)
Hybrid Electric	-0.024	0.121	0.157	0.064	0.048
	(0.121)	(0.087)	(0.103)	(0.070)	(0.096)
Hydrogen Fuel Cell	-0.100	0.062	0.082	0.005	0.017
	(0.142)	(0.081)	(0.099)	(0.093)	(0.138)
Plug-in Hybrid Electric	-0.154	-0.071	-0.072	-0.347***	-0.146
	(0.115)	(0.099)	(0.104)	(0.110)	(0.124)
Battery Electric	-0.244**	-0.019	-0.072	-0.005	-0.047
	(0.109)	(0.099)	(0.121)	(0.089)	(0.134)
Fuel cost	-0.032***	-0.026***	-0.032***	-0.034***	-0.039***
	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)
CO2 Emissions	-0.152**	-0.205***	-0.244***	-0.414***	-0.253***
	(0.074)	(0.066)	(0.066)	(0.127)	(0.063)
Recharge time	-0.040**	-0.030**	-0.039**	-0.034*	-0.032
	(0.016)	(0.012)	(0.017)	(0.021)	(0.020)
Refill availability	0.293***	0.244***	0.335***	0.317***	0.411***
	(0.083)	(0.061)	(0.076)	(0.101)	(0.088)
Driving range	0.027***	0.034***	0.038***	0.043***	0.041***
	(0.010)	(0.008)	(0.010)	(0.007)	(0.007)
-1 * Price	0.175***	0.145***	0.171***	0.244***	0.238***
	(0.017)	(0.021)	(0.028)	(0.026)	(0.025)
ϕ		1.249***	2.244***	3.117***	3.119***
		(0.107)	(0.249)	(0.342)	(0.329)
τ				1.109***	0.696***
				(0.0851)	(0.0794)
γ					-0.816***
					(0.193)
No. of parameters	14	14	27	28	29
No. of choice sets	1612	1612	1612	1612	1612
LL		538.3	653.4	762.6	791.4
AIC		-1048.6	-1252.8	-1469.3	-1524.7
BIC		-973.2	-1107.4	-1318.5	-1368.6

Notes: This table reports population mean parameters. Table C1 of Appendix A reports corresponding standard deviations. Standard errors in parentheses are clustered at the respondent level (202 respondents). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

significant at the 5% level in models that do not account for scale heterogeneity (LAD, F-MNL, and F-NMXL). Of note is the disutility associated with LPG vehicles, which complements well the declining popularity of this fuel type since our survey was conducted Laverick (2023). However, the coefficient is only significant (1% level) in the F-GMNL specifications, demonstrating the value of these more flexible specifications in uncovering preference relationships that may be hidden in simpler specifications that do not account for variable individual perceptions of scenario incompleteness. Vehicle size is not a robust independent predictor of choice expectations, but as we will discuss next, there is considerable preference heterogeneity for this attribute.

For the F-NMXL, FGMNL-II and F-GMNL models, we also estimate the population standard deviations of $\boldsymbol{\beta}_n$ and λ_n around the average coefficients.[27] These estimates are presented in Table C1 of Appendix C. Across all three fractional response models, we find significant heterogeneity in consumer preferences for price, fuel cost, driving range, recharge time and CO2 emissions, in addition to several of the qualitative dummies, including the preferences for battery electric and large size vehicles. We will return to quantify this preference variability in the next subsection.

The estimated scale parameter τ is 1.11 in F-GMNL-II with a standard error of 0.09, which suggests the presence of substantial scale heterogeneity in the data. The F-GMNL estimate of the scale parameter τ is 0.70 with a standard error of 0.08. Meanwhile, the F-GMNL estimate of γ is -0.816, which implies that the variance of residual taste heterogeneity increases more than proportionately as the vector of utility weights are scaled up or down by σ_n (Keane and Wasi, 2013). Recall the distinction between the scale parameter of the random utility model (σ) and the scale parameter of the beta regression (ϕ). The former inversely relates to an individual's uncertainty of preference (*e.g.*, due to the perception of incompleteness concerning a choice scenario), whereas the latter represents the behavioral error in stating that preference during the survey (*e.g.*, rounding behavior). Our estimates of the behavioral error parameter ϕ are significantly greater than zero in all of the fractional response models (1% level). This highlights the importance of capturing this form of measurement error—separate to coefficient and scale heterogeneity—when estimating preferences from elicited choice probabilities.

Finally, we use the Akaike information criterion (AIC) and the Bayesian infor-

charges, may provide a strong economic rationale for preferring vehicles with lower emissions.

[27] We assume that each random coefficient is independently normal. This assumption rules out correlations in tastes for different attributes. Relaxing this assumption and estimating an unrestricted covariance matrix would entail the estimation of a further 105 parameters, which is impractical to implement in the present application.

mation criterion (BIC) to compare the fit of the different models in our dataset. Adding residual taste heterogeneity (F-NMXL) leads to a marked improvement in the likelihood over F-MNL, from 538.3 to 653.4 (*i.e.*, 115.1 points or 21%), with corresponding improvements in both AIC and BIC. Adding scale heterogeneity (F-GMNL-II) adds just one more parameter and leads to a 16.7% improvement in the likelihood over F-NMXL of another 109.2 points, resulting in a further large improvement in the AIC and BIC. This aspect of our choice probabilities data is aligned with the results of Fiebig et al. (2010) and Keane and Wasi (2013) based on traditional (*i.e.*, "pick one") choice data, which suggest that scale heterogeneity accounts for a substantial share of interpersonal heterogeneity in choice behavior. We also observe a small improvement in model fit when moving from F-GMNL-II to F-GMNL.

4.3 Willingness-to-Pay (WTP) for Vehicle Attributes

The coefficient estimates reported in the previous subsection express preferences in the abstract unit of *utils*, making it difficult to evaluate their fuller quantitative implications. To convert these estimates into a more natural unit of measurement, a standard approach is to study WTP implied by the model, derived as the ratio of the coefficient on the respective non-price attribute to the coefficient on the price attribute (*e.g.*, Blass et al., 2010; Wiswall and Zafar, 2018; Koşar et al., 2022).

In Table 4, we evaluate these WTP measures at the average preference coefficients, by applying the formula $\bar{\boldsymbol{\beta}}/\bar{\lambda}$. Since we have coded our price attribute in 10,000s of Australian dollars, the unit for the WTP estimates is also AUD $10,000s. Given our coding scheme for the non-price attributes, the reported numbers correspond to the WTP for each of the following changes: (i) $1 increase in fuel cost per 100 kilometres; (ii) one-hour increase in recharge time (iii) 100 kilometre increase in driving range; (iv) 100% increase in CO2 emissions (percent of average vehicle); (v) 100% increase in refill availability (percent of service stations); (vii) switching from small to medium or large vehicle size; and (vii) switching from conventional to one of the five alternative fuel types.

The results largely agree with our discussion of the preference parameter estimates. Depending on the model specification, the average consumer's WTP is estimated at $1,390 to $1,860 for a $1 reduction in fuel cost per 100 kilometres; $1,330 to $2,290 for a one-hour reduction in recharge time; $1,540 to $2,350 for a 100 kilometre increase in driving range; $869 to $1,694 for a 10% reduction in CO2 emissions; and $1,296 to $1,954 for a 10% increase in refill availability. The majority of the WTP estimates for the vehicle size and fuel type dummies are not

Table 4: WTP estimates for AFVs derived from estimated population preference parameters using choice probabilities (AUD $10,000s).

	LAD	F-MNL	F-NMXL	F-GMNL-II	F-GMNL
Medium size	0.384	0.522	0.400	0.616***	0.336
	(0.401)	(0.357)	(0.425)	(0.220)	(0.250)
Large size	0.457	0.852**	0.641	0.219	0.148
	(0.395)	(0.341)	(0.455)	(0.347)	(0.281)
Liquid Petroleum Gas	-0.739	-0.550	-0.571	-1.932***	-1.298***
	(0.628)	(0.518)	(0.570)	(0.346)	(0.469)
Hybrid Electric	-0.140	0.833	0.916	0.262	0.204
	(0.696)	(0.570)	(0.561)	(0.283)	(0.395)
Hydrogen Fuel Cell	-0.570	0.429	0.476	0.019	0.072
	(0.825)	(0.540)	(0.574)	(0.380)	(0.579)
Plug-in Hybrid Electric	-0.882	-0.489	-0.422	-1.419***	-0.612
	(0.672)	(0.694)	(0.617)	(0.416)	(0.547)
Battery Electric	-1.396**	-0.131	-0.422	-0.019	-0.198
	(0.659)	(0.686)	(0.711)	(0.364)	(0.572)
Fuel cost	-0.186***	-0.179***	-0.186***	-0.139***	-0.164***
	(0.036)	(0.036)	(0.045)	(0.021)	(0.029)
CO2 Emissions	-0.869**	-1.418***	-1.425***	-1.694***	-1.063***
	(0.429)	(0.435)	(0.428)	(0.446)	(0.281)
Recharge time	-0.227**	-0.208**	-0.229*	-0.140	-0.133
	(0.096)	(0.089)	(0.118)	(0.088)	(0.082)
Refill availability	1.676***	1.682***	1.954***	1.296***	1.725***
	(0.498)	(0.481)	(0.478)	(0.452)	(0.395)
Driving range	0.154***	0.235***	0.219***	0.178***	0.173***
	(0.059)	(0.076)	(0.082)	(0.035)	(0.031)

Notes: WTP figures respectively correspond to the following units: vehicle size (vs. small), fuel type (vs. conventional), $1 increase in fuel cost per 100km, 100% increase in CO2 emissions (% of average vehicle), 1 hour increase in recharge time, 100% increase in refill availability (% of service stations), 100km increase in driving range. Standard errors in parentheses are clustered at the respondent level (202 clusters). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

significantly different from zero in a consistent manner across models. Nevertheless, for LPG vehicles, we do find significant WTP estimates of $-$\$1,298 (F-GMNL) and $-$\$1,932 (F-GMNL-II) in the two models that account for scale heterogeneity on top of coefficient heterogeneity.

As discussed in the context of the LAD estimation of equation (7), a key drawback of the ratio $\bar{\boldsymbol{\beta}}/\bar{\lambda}$ is that it does not correspond to a usual central tendency measure (*e.g.*, mean or median) of the WTP distribution *per se*. In the case of fractional response models for which we estimate the population distribution of each

Table 5: Population WTP distribution (AUD \$10,000s), simulated using 10,000 draws of preference parameters from F-NMXL.

Percentile	5th	10th	25th	50th	75th	90th	95th	$\bar{\beta}/\bar{\lambda}$
Medium size	-2.905	-1.438	-0.404	0.111	0.681	1.770	3.513	0.400
Large size	-2.878	-1.208	-0.072	0.327	0.793	1.844	3.524	0.641
Liquid Petroleum Gas	-6.084	-3.277	-1.380	-0.477	0.327	2.206	4.825	-0.571
Hybrid Electric	-1.737	-0.766	0.136	0.251	0.481	1.038	2.004	0.916
Hydrogen Fuel Cell	-2.691	-1.376	-0.532	-0.110	0.293	1.084	2.339	0.476
Plug-in Hybrid Electric	-11.11	-5.668	-2.239	-0.475	1.153	4.597	10.44	-0.422
Battery Electric	-13.06	-6.632	-2.380	-0.257	1.807	5.718	12.04	-0.422
Fuel cost	-1.242	-0.672	-0.278	-0.101	0.050	0.423	1.013	-0.186
CO2 Emissions	-10.63	-5.499	-2.210	-0.602	0.862	3.876	8.424	-1.425
Recharge time	-1.020	-0.537	-0.234	-0.091	0.024	0.340	0.806	-0.229
Refill availability	-9.162	-3.936	-0.552	0.938	2.762	6.478	11.91	1.954
Driving range	-1.674	-0.713	-0.116	0.142	0.455	1.093	2.044	0.219

Notes: For variable definitions, see notes to Table 4. The simulation is based on the population distribution of preference parameters characterized by the mean (Table 3) and standard deviation (Table C1 in Appendix C) estimates for the F-NMXL model. The final column ($\bar{\beta}/\bar{\lambda}$) reproduces the average consumer's WTP reported in the F-NMXL column of Table 4, to facilitate comparisons with the simulated percentiles of β_n/λ_n.

preference parameter on top of their averages, we are able to make draws of $\boldsymbol{\beta}_n$ and λ_n from their estimated distributions. The results therefore allow us to derive and examine the implied WTP distributions concerning $\boldsymbol{\beta}_n/\lambda_n$.

To demonstrate, in Table 5 we present the simulated distribution of WTPs based on the F-NMXL preference estimates in Table 3. To facilitate comparisons with the percentiles of this distribution, we also reproduce the average consumer's WTPs (*i.e.*, $\bar{\boldsymbol{\beta}}/\bar{\lambda}$) reported in the F-NMXL column of the preceding table. The mean WTP does not possess finite moments in this model, but the median WTP is well-defined. In general, differences between the median WTP and the average consumer's WTP within the F-NMXL model tends to be larger than differences between the LAD and F-NMXL estimates of the average consumer's WTP. For example, consider again the average consumer's WTP for a 10% increase in refill availability, which was estimated at \$1,676 (LAD) or \$1,954 (F-NMXL). The median of the WTP distribution under F-NMXL, however, occurs at a much lower amount of \$938, which underscores the potential importance of examining the WTP distribution directly. Nevertheless, quantifying the WTP distribution in this manner has its limitations, as draws of λ_n near zero result in implausibly large magnitudes of WTP values. In Table 5, this issue is evident from the several-fold increases in percentile values observed when moving to the left of the 25th percentile or to the right of the 75th percentile.

Table 6: Population WTP distribution (AUD $10,000s), simulated using 10,000 draws of WTP parameters from F-NWTP.

Percentile	5th	10th	25th	50th	75th	90th	95th
Medium size	-1.220	-0.872	-0.301	0.330	0.972	1.579	1.920
Large size	-1.108	-0.790	-0.273	0.307	0.894	1.414	1.697
Liquid Petroleum Gas	-2.634	-2.279	-1.648	-0.983	-0.290	0.318	0.678
Hybrid Electric	-1.432	-1.143	-0.649	-0.094	0.446	0.948	1.244
Hydrogen Fuel Cell	-2.889	-2.249	-1.141	0.072	1.273	2.347	2.983
Plug-in Hybrid Electric	-4.188	-3.523	-2.429	-1.208	-0.009	1.070	1.692
Battery Electric	-5.045	-4.094	-2.428	-0.513	1.345	2.994	3.949
Fuel cost	-0.474	-0.408	-0.300	-0.178	-0.056	0.054	0.118
CO2 Emissions	-5.886	-4.639	-2.640	-0.459	1.752	3.743	4.951
Recharge time	-0.764	-0.660	-0.482	-0.286	-0.091	0.085	0.184
Refill availability	-2.090	-1.198	0.282	1.943	3.594	5.062	5.950
Driving range	-0.365	-0.245	-0.051	0.172	0.391	0.586	0.703

Notes: For variable definitions, see notes to Table 4. The simulation is based on the population distribution of WTP parameters characterized by the mean and standard deviation estimates for the F-NWTP model. The F-NWTP estimates are reported in Appendix C (Table C2). Since the model assumes normally distributed WTP parameters, the average consumer's WTP coincides with the median (50th percentile).

The NWTP model of Train and Weeks (2005), presented in equation (4), addresses these limitations directly by parameterizing and estimating the model in terms of the WTP distribution rather than the utility parameter distribution. In Table C2 of Appendix C, we report a fractional response model which incorporates the NWTP preference structure (F-NWTP).[28] In Table 6, we present the simulated distributions of population-level heterogeneity in WTPs based on the F-NWTP model. Consistent with findings in studies estimating the NWTP preference structure from stated choices, we observe that the tails of the distribution are pulled in towards the median compared to those in NMXL. The median WTP coincides with the mean WTP in this model, as the population distribution of each WTP measure is specified to be normal. The issue of whether the mean WTP in the F-NWTP model aligns more closely with the average consumer's WTP or the median WTP in F-NMXL calls for a case-by-case assessment. However, in many instances, the mean falls within or slightly outside an interval between the two estimates, with relative proximity to each end of the interval varying from one attribute to another.

[28]The same table also presents a special case of F-NWTP which assumes homogeneous WTP coefficients for all individuals, which we dub F-WTP. The F-WTP estimates are simply provided to facilitate comparisons; in terms of substance, they are numerically identical to the WTP values implied by the F-MNL column in Table 3 by construction. As explained by Train and Weeks (2005), whether one specifies the model in the WTP space or utility space is irrelevant as the model does not involve random coefficients to represent interpersonal preference heterogeneity.

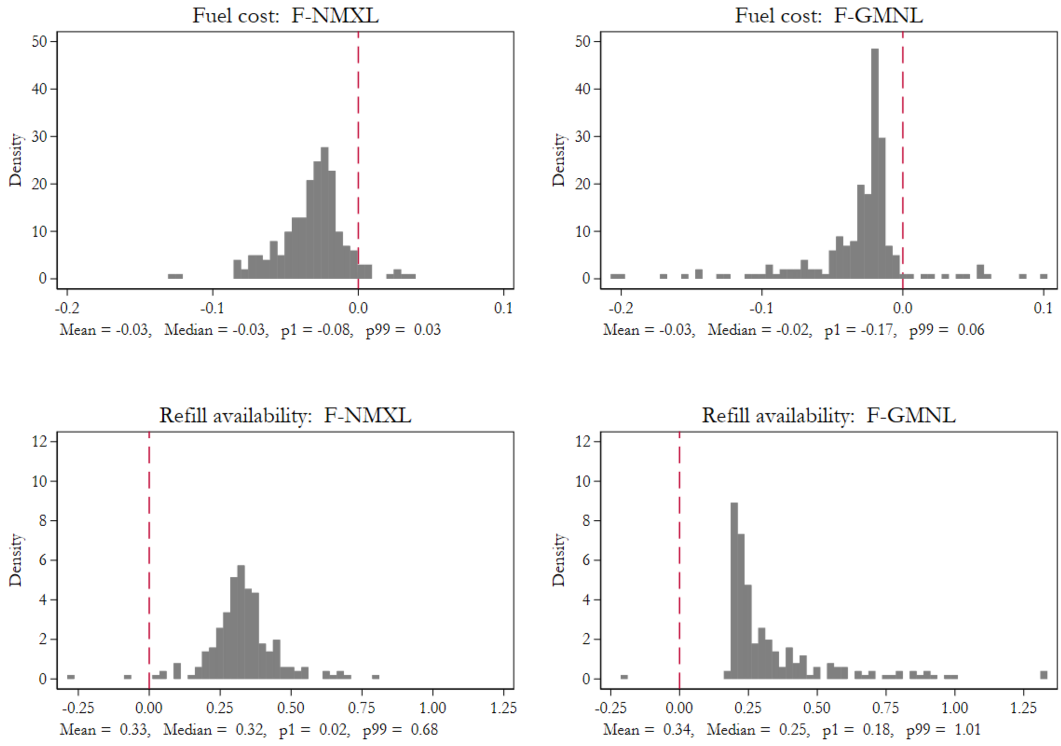

Figure 2: Posterior distributions of individual-level preference parameters for the choice attributes *Fuel cost* and *Refill availability* in F-NMXL (column 1) and F-GMNL (column 2). Posterior distributions for the remaining numeric choice attributes and for the F-GMNL-II model are reported in Appendix C (Figures C1 and C2).

4.4 Individual-level Preferences

In this subsection, we evaluate individual-level parameters implied by the fractional response models of preference heterogeneity. The results will help us to better understand the behavioral reasons why scale heterogeneity, in addition to coefficient heterogeneity, improves model fit in the data (cf. Table 3). These parameters take the form of individual-specific posterior means, derived by applying Bayes' rule to combine information available from the estimated population distributions, which act as priors, with information from a given individual's stated responses to the DCE. Train (2009, §11) offers an accessible guide to this approach. In the case of the F-NMXL model, each prior is a normal distribution. In the cases of F-GMNL-II and F-GMNL, we evaluate the aggregated effects of scale and coefficient heterogeneity (*e.g.*, posterior mean of $\sigma_n \boldsymbol{\beta}_n$ in F-GMNL-II rather than that of $\boldsymbol{\beta}_n$ alone). As Fiebig et al. (2010) and Keane and Wasi (2013) point out, each prior in this case may be interpreted as a more flexible distribution characterizable as a continuous mixture of scaled normals.

To demonstrate the implications of modelling scale heterogeneity in addition to coefficient heterogeneity, Figure 2 summarizes the distributions of the individual-level parameters for the fuel cost and refill availability choice attributes in the F-NMXL and F-GMNL models. The posterior distributions for the remaining four numeric choice attributes and all corresponding distributions for the F-GMNL-II model are reported in Appendix C (Figures C1 and C2). In general, the distributions for the F-NMXL model tend to be more bell-shaped and so are not well-suited to accommodating consumers with lexicographic preferences, in a broad sense of placing great weight on a single choice attribute. This is because the normal prior is less flexible for capturing outliers in the dataset. The distributions of the individual-level parameters in the F-GMNL and F-GMNL-II models depart substantially from normality and can account for the pronounced left or right skewness in the data. Indeed, this is observed for preferences over the fuel cost and refill availability choice attributes in Figure 2. This flexibility is important for variables where we simultaneously observe individuals who have strong preferences over a particular attribute and a distinct group of individuals who exhibit a low responsiveness to variation in this attribute. Incorporating scale heterogeneity allows for such a pattern to emerge by allocating a small scale parameter value to the former group of individuals and a large scale parameter value to the latter group of individuals. The ability to address this source of heterogeneity is an advantage that the fractional response model has over the LAD estimation, as we have explained in the context of equation (8).

4.5 Comparison with Standard DCE Using Choice Format

We have provided evidence of substantial choice uncertainty in our dataset. This observation alone suggests that deterministic choices, which standard DCEs elicit, would be inherently restrictive in the AFV setting and points to the value of eliciting choice probabilities. As explained in Section 3, participants in our study were randomly allocated to either the probability elicitation format, the data from which we have been analyzing so far, or the standard pick-one choice format. We now study data on the latter group of 525 respondents who submitted a total of 4,200 conditional choices (eight choices each). This gives us an opportunity to evaluate how different our results would be with the standard elicitation format.

Before turning to the results, we note that in general, empirical identification of preference structures featuring scale heterogeneity (that is, NWTP, GMNL-II, and GMNL) is fragile when we use the stated choice data, suggesting that allowing respondents to express their uncertainty in the form of choice probabilities facilitates estimation of rich behavioral models in the AFV settings that would otherwise not be estimable. We employed computational compromises to estimate these models using the stated choice data that we had not applied earlier and despite the larger sample size. Specifically, for NWTP and GMNL-II, we had to assume homogeneous preferences for the vehicle size and fuel type attributes, and only allow for heterogeneity concerning the numerical attributes. In the case of the full GMNL model, identification is more dubious as we had to set a small number of Halton draws (50) to accomplish numerical convergence.[29]

With these caveats in mind, Table 7 presents the population level WTP at mean estimates, $\bar{\beta}/\bar{\lambda}$, for the stated choice data using the MNL, NMXL, GMNL-II and GMNL models.[30] A direct comparison with Table 4 tells us that the WTP estimates for the numeric attributes remain largely unchanged in terms of signs and significance, although they exhibit different magnitudes which can be important from a non-market valuation perspective. Concerning the vehicle size dummies, we obtain positive and significant (p-values < 0.01) estimates in all four models of stated choice data, in contrast to our earlier results based on choice probabilities data where they were insignificant at the 10% level in four out of five specifications. One possible explanation for this sensitivity is that the size of vehicle required in the future depends on factors such as family size and lifestyle needs, which may entail a relatively high level of uncertainty. We would also expect preferences for vehicle size to be highly heterogeneous for these reasons; however, evidence that preferences for these attributes are heterogeneous is less equivocal in the stated choice data (Table C4) than in the choice probabilities data (Table C1).

Additionally, the battery electric dummy variable is negative and significant at the 5% level in the stated choice data, but negative and insignificant at the 10% level in four out of five specifications for the choice probabilities data. The future market prospect for battery electric vehicles, which might affect factors such as expected resale price, was highly uncertain at the time of the survey implementation in 2017 (and remains uncertain today).

The above results are aligned with what we observe in the WTP space models for the stated choice data (see Table C5 in Appendix C).

[29] The exact value of a simulated likelihood function is sensitive to the number of Halton draws used. Therefore, we cannot directly compare model fit where there is a varied number of Halton draws.

[30] The corresponding population preference parameter mean and standard deviation estimates are reported in Appendix C (Tables C3 and C4).

Table 7: WTP estimates for AFVs derived from estimated population preference parameters using stated choices (AUD $10,000s).

	(1) MNL	(2) NMXL	(3) GMNL-II	(4) GMNL
Medium size	1.263***	0.784***	0.716***	0.749***
	(0.226)	(0.179)	(0.166)	(0.201)
Large size	1.278***	0.711***	0.675***	0.743***
	(0.222)	(0.188)	(0.114)	(0.201)
Liquid Petroleum Gas	-0.057	-0.440	-0.450***	-0.444
	(0.380)	(0.350)	(0.168)	(0.396)
Hybrid Electric	0.221	-0.046	0.118	0.195
	(0.385)	(0.320)	(0.217)	(0.376)
Hydrogen Fuel Cell	-0.213	-0.246	-0.104	-0.223
	(0.375)	(0.319)	(0.237)	(0.332)
Plug-in Hybrid Electric	-0.621	-0.803**	-0.537***	-0.655*
	(0.383)	(0.327)	(0.189)	(0.398)
Battery Electric	-0.597*	-0.778**	-0.682***	-0.732**
	(0.346)	(0.310)	(0.209)	(0.350)
Fuel cost	-0.189***	-0.168***	-0.182***	-0.188***
	(0.019)	(0.017)	(0.020)	(0.020)
CO2 Emissions	-1.133***	-0.969***	-1.254***	-1.149***
	(0.266)	(0.221)	(0.268)	(0.254)
Recharge time	-0.362***	-0.281***	-0.350***	-0.346***
	(0.058)	(0.051)	(0.047)	(0.064)
Refill availability	2.497***	1.949***	1.805***	1.983***
	(0.319)	(0.269)	(0.281)	(0.285)
Driving range	0.344***	0.268***	0.269***	0.292***
	(0.039)	(0.035)	(0.032)	(0.035)

Notes: For variable definitions, see notes to Table 4. Standard errors in parentheses are clustered at the respondent level (525 respondents). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.6 Comparisons with Maximum Score (MS) Estimator

The MS estimator (Manski, 1975, 1985) offers an attractive alternative to the LAD estimator when one's interest lies in the semi-parametric estimation of the average consumer's WTP coefficients, $\bar{\boldsymbol{\beta}}/\bar{\lambda}$. As summarized in Section 2.4, the MS estimator can accommodate more flexible model specifications that feature any form of scale heterogeneity and non-logit forms of error terms. Moreover, it can be applied to both choice probabilities and stated choice data, enabling us to draw semi-parametric comparisons between the two data sources.

Table 8 presents the MS estimates of $\bar{\boldsymbol{\beta}}/\bar{\lambda}$ for the choice probabilities data, and

Table 9 presents corresponding results for the standard choice data. Since the objective function is a step function, following Fox (2007) and Yan and Yoo (2019), we compute the parameter estimates by applying differential evolution (DE) algorithms rather than usual numerical maximization methods. A main drawback of the MS estimator is that it is only interval-identified, meaning that different sets of parameter values may result in the same maximum. In light of this property, we restart the DE algorithms 1,000 times for the choice probabilities data, and report descriptive statistics on multiple solutions found therefrom. Executing the 1,000 restarts took us approximately 10 days of computer run time.[31] The standard choice data includes more than 2.5 times as many choice sets (4,200 versus 1,612) and each restart took approximately twice as much run time. For practical reasons, we therefore limit the number of restarts for the standard choice data to 500.

We first consider the results in Table 8 for the choice probabilities data. We observe that the the MS estimates are often more aligned with the fractional response model estimates than with the LAD estimates. We previously found the five numeric attributes to be robust predictors of WTP (see Table 4). The new MS estimate concerning fuel cost occurs in the short interval $[-0.136, -0.123]$; for comparison, the closest point estimate found previously was in F-GMNL-II (-0.139). The MS estimate for CO2 emissions occurs in the interval $[-0.220, -0.135]$, which spans a far smaller range of values (in absolute terms) than the point estimates found previously, even compared to the closest estimate which is based on LAD (-0.869). For recharge time, the MS estimate occurs in the interval $[-0.344, -0.307]$; the closest point estimates found previously were in LAD (-0.227) and F-NMXL (-0.229). The MS estimate concerning refill availability occurs in the interval $[1.810, 2.385]$ and the F-NMXL model estimate was contained in this interval (1.954). Finally, the MS estimate concerning driving range occurs in the interval $[-0.111, -0.081]$; in contrast, all previous estimates indicated that the average consumer's WTP for driving range was positive (although, as we observed in Section 4.3, this masks substantial heterogeneity in the population distribution).

We next consider the results in Table 9 for the stated choices data. We previously found the size dummies and the five numeric attributes to be robust predictors of WTP (see Table 7). We now find that the MS estimates occur in the intervals $[0.272, 0.666]$ for the medium size dummy and $[0.492, 1.094]$ for the large size dummy. The corresponding point estimates for the models with preference hetero-

[31]Our estimation was executed on a Windows workstation, using Stata 17/MP to utilize 6 CPU cores for parallel computing. We applied the usual computational configuration of the DE algorithms. The *population size* was set to 40 times the number of coefficients (that is, set to 480 in our application), and the *number of generations* was set to 10 times the population size. The *cross-over probability* was set to 0.8 and the *amplification factor* to 0.4.

Table 8: Maximum score estimation of WTP using choice probabilities (AUD $10,000s).

	Mean	SD	Min	Max	IQR
Medium size	-0.579	0.005	-0.583	-0.576	0.007
Large size	0.615	0.314	0.393	0.837	0.444
Liquid Petroleum Gas	3.560	0.055	3.521	3.598	0.078
Hybrid Electric	2.252	0.380	1.984	2.521	0.537
Hydrogen Fuel Cell	0.754	0.265	0.567	0.941	0.374
Plug-in Hybrid Electric	1.412	0.244	1.239	1.584	0.345
Battery Electric	0.195	0.234	0.029	0.361	0.332
Fuel cost	-0.130	0.009	-0.136	-0.123	0.013
CO2 Emissions	-0.177	0.061	-0.220	-0.135	0.086
Recharge time	-0.326	0.026	-0.344	-0.307	0.037
Refill availability	2.097	0.407	1.810	2.385	0.575
Driving range	-0.096	0.022	-0.111	-0.081	0.031
Mean score	0.720				
No. of choice sets	1612				

Notes: For variable definitions, see notes to Table 4.

Table 9: Maximum score estimation of WTP using stated choices (AUD $10,000s).

	Mean	SD	Min	Max	IQR
Medium size	0.468	0.101	0.272	0.666	0.118
Large size	0.734	0.164	0.492	1.094	0.187
Liquid Petroleum Gas	-0.756	0.404	-1.398	0.078	0.573
Hybrid Electric	-3.660	0.280	-4.112	-3.249	0.549
Hydrogen Fuel Cell	-3.035	0.456	-3.863	-2.366	0.552
Plug-in Hybrid Electric	-3.454	0.268	-4.121	-3.188	0.353
Battery Electric	-2.966	0.272	-3.516	-2.626	0.360
Fuel cost	-0.068	0.006	-0.080	-0.058	0.008
CO2 Emissions	-0.064	0.216	-0.451	0.270	0.324
Recharge time	-0.132	0.033	-0.169	-0.065	0.027
Refill availability	1.540	0.348	0.981	2.225	0.479
Driving range	0.247	0.022	0.190	0.280	0.024
Mean score	0.731				
No. of choice sets	4200				

Notes: For variable definitions, see notes to Table 4.

geneity (NMXL, GMNL-II and GMNL) are all contained in the large size interval and moderately to the right of the medium size interval; the MNL estimates are somewhat larger. We find that the MS estimate concerning fuel cost occurs in the short interval $[-0.080, -0.058]$. Our previous point estimates were found moderately to the left of this interval. The MS estimate concerning CO2 emissions occurs in the long interval $[-0.451, 0.270]$, which, as with the choice probabilities data, spans values to the right of the point estimates found previously. For recharge time, the MS estimate occurs in the interval $[-0.169, -0.065]$; the closest point estimate found previously was in the F-NMXL model (-0.281). Finally, the MS estimates of refill availability and driving range occur in the intervals $[0.981, 2.225]$ and $[0.190, 0.280]$, respectively. The previous estimates for the models with preference heterogeneity were contained in (or just outside) these intervals.

Turning to the WTP for fuel types—Liquid Petroleum Gas through Battery Electric)—we observe sign reversals across the board between the two datasets. In many instances, the MS estimates also yield implausibly large values, surpassing AUD $30,000 in magnitude. Our analyses above have indicated that the WTP for these attributes is generally insignificant, using a less flexible semi-parametric method (LAD) for the probabilities data, as well as parametric methods for both types of data. The new MS results suggest that it is particularly challenging to obtain stable estimates concerning such attributes under more flexible semi-parametric assumptions. This contextualization of the MS results further highlights the complementary advantages of applying both semi-parametric and parametric methods in non-market valuation studies.

We conclude by comparing the overall prediction success rates between the two data sets. For choice probabilities, a correct prediction occurs if the MS estimator of the utility index favors the option that attracted 50% or higher probabilities. For stated choices, a correct prediction occurs if the MS estimator of the utility index favors the actual choice. Since the score for an observed response is equal to 1 in those scenarios where the prediction is correct in these senses and 0 otherwise, the MS estimator can be seen as an estimator that maximizes the prediction success rate for each type of data. The maximized sample mean scores indicate that our MS estimates can correctly predict responses on 72% of scenarios in the choice probabilities data and on 73.1% of scenarios in the standard choice data.

5 Conclusions

The inherent parsimony of DCE designs often results in choice scenarios that may seem incomplete from the respondent's perspective. In such settings, asking respondents to state their choice expectations, in the form of probabilities with which they would make particular decisions, is a more natural elicitation format than asking them to state their preferred choices. The current method for analyzing elicited choice probabilities primarily employs semi-parametric estimation of population average preferences within the mixed logit framework, without fully specifying preference distributions. However, this flexibility in distributional assumptions presents challenges in drawing inferences about preference heterogeneity, which is a central focus of research in various fields of applied microeconomics.

To complement the semi-parametric method, we introduce a fractional response model based on a mixture of beta distributions. This model enables researchers to uncover preference heterogeneity under the same set of parametric assumptions concerning preference heterogeneity as those used in the analysis of elicited choices. In particular, the model can accommodate multi-faceted preference structures which cannot be consistently estimated by the semi-parametric method (*e.g.*, GMNL and mixed logit in the WTP space, which exhibit both coefficient and scale heterogeneity). We present an empirical analysis using data from a DCE on alternative fuel vehicles, illustrating the complementary roles of the two approaches. For simple preference structures, to which the semi-parametric method is robust, our findings show a close alignment between semi-parametric estimates and fractional response model estimates. Nevertheless, all specifications of the fractional model indicate substantial preference heterogeneity, highlighting the importance of considering more than average preferences. Additionally, our analysis of a companion DCE with the traditional choice elicitation format suggests that identifying multi-faceted preference structures is more computationally tenable with elicited probabilities, further emphasizing the benefits of this elicitation format.

Eliciting subjective choice expectations is relatively new in DCE research, yet it has already led to a small but influential body of empirical studies (*e.g.*, Blass et al., 2010; Wiswall and Zafar, 2018; Koşar et al., 2022). We are optimistic that our contribution will enable researchers to more fully exploit this elicitation format, by reducing the analytic cost associated with this approach, which often involves bypassing the analysis of preference heterogeneity.

References

Abdoolakhan, Z. (2010). *Acceptance of alternative fuel and hybrid vehicles in Australia: results based on survey data, choice modelling and elasticity estimation*. Ph. D. thesis, University of Western Australia.

Arcidiacono, P., V. J. Hotz, and S. Kang (2012). Modeling college major choices using elicited measures of expectations and counterfactuals. *Journal of Econometrics 166*(1), 3–16. Annals Issue on "Identification and Decisions", in Honor of Chuck Manski's 60th Birthday.

Australian Automobile Association (2023). EV Index. https://data.aaa.asn.au/ev-index/. Accessed 2023-12-12.

Australian Bureau of Statistics (2018). Motor Vehicle Census (Cat. 9309.0). https://www.abs.gov.au/ausstats/abs@.nsf/allprimarymainfeatures/-77C32540D6631DE7CA2584430019F9EC?opendocument. Accessed 2023-13-12.

Beck, M. J., J. M. Rose, and D. A. Hensher (2013). Environmental attitudes and emissions charging: An example of policy implications for vehicle choice. *Transportation Research Part A: Policy and Practice 50*, 171–182.

Beggs, S., S. Cardell, and J. Hausman (1981). Assessing the potential demand for electric cars. *Journal of Econometrics 17*(1), 1–19.

Blass, A. A., S. Lach, and C. F. Manski (2010). Using elicited choice probabilities to estimate random utility models: Preferences for electricity reliability. *International Economic Review 51*(2), 421–440.

Bleemer, Z. and B. Zafar (2018). Intended college attendance: Evidence from an experiment on college returns and costs. *Journal of Public Economics 157*, 184–211.

Boyer, M., P. De Donder, C. Fluet, M.-L. Leroux, and P.-C. Michaud (2017). Long-term care insurance: Knowledge barriers, risk perception and adverse selection. Working Paper 23918, National Bureau of Economic Research.

Daly, A., S. Hess, and K. Train (2012). Assuring finite moments for willingness to pay in random coefficient models. *Transportation 39*, 19–31.

Delavande, A. (2008). Pill, patch, or shot? subjective expectations and birth control choice. *International Economic Review 49*(3), 999–1042.

Delavande, A. and C. F. Manski (2015). Using elicited choice probabilities in hypothetical elections to study decisions to vote. *Electoral Studies 38*, 28–37.

Delavande, A. and B. Zafar (2019). University choice: The role of expected earnings, nonpecuniary outcomes, and financial constraints. *Journal of Political Economy 127*(5), 2343–2393.

Dominitz, J. and C. F. Manski (1997). Using expectations data to study subjective income expectations. *Journal of the American Statistical Association 92*(439), 855–867.

Electric Vehicle Council (2023). State of Electric Vehicles Report 2023. https://electricvehiclecouncil.com.au/reports/soevs-report-2023/.

Ferrari, S. and F. Cribari-Neto (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics 31*(7), 799–815.

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 29*(3), 393–421.

Fox, J. T. (2007). Semiparametric estimation of multinomial discrete-choice models using a subset of choices. *RAND Journal of Economics 38*(4), 1002–1019.

Hackbarth, A. and R. Madlener (2013). Consumer preferences for alternative fuel vehicles: A discrete choice analysis. *Transportation Research Part D: Transport and Environment 25*, 5–17.

Hackbarth, A. and R. Madlener (2016). Willingness-to-pay for alternative fuel vehicle characteristics: A stated choice study for germany. *Transportation Research Part A: Policy and Practice 85*, 89–111.

Hendren, N. (2013). Private information and insurance rejections. *Econometrica 81*(5), 1713–1762.

Hensher, D., J. Louviere, and J. Swait (1998). Combining sources of preference data. *Journal of Econometrics 89*(1-2), 197–221.

Herriges, J., S. Bhattacharjee, and C. Kling (2011). Capturing preferences under incomplete scenarios using elicited choice probabilities. Working paper no. 11003 march 2011, Iowa State University, Department of Economics, Ames, Iowa.

Hurd, M. D., J. P. Smith, and J. M. Zissimopoulos (2004). The effects of subjective survival on retirement and social security claiming. *Journal of Applied Econometrics 19*(6), 761–775.

Juster, F. T. (1966). Consumer buying intentions and purchase probability: An experiment in survey design. *Journal of the American Statistical Association 61*(315), 658–696.

Keane, M. and N. Wasi (2013). Comparing alternative models of heterogeneity in consumer choice behavior. *Journal of Applied Econometrics 28*(6), 1018–1045.

Koşar, G. and C. O'Dea (2023). Chapter 21 - Expectations data in structural microeconomic models. In R. Bachmann, G. Topa, and W. van der Klaauw (Eds.), *Handbook of Economic Expectations*, pp. 647–675. Academic Press.

Koşar, G., T. Ransom, and W. van der Klaauw (2022). Understanding migration aversion using elicited counterfactual choice probabilities. *Journal of Econometrics 231*(1), 123–147.

Laverick, P. (2023). As LPG vehicles disappear from Australia's roads, remaining drivers struggle to find fuel. ABC News 19 April 2023, https://www.abc.net.au/news/2023-04-19/lpg-cars-disappearing-from-roads-gas-renewable-fuel-vehicles/102236128. Accessed 2023-12-12.

Layton, D. F. and G. Brown (2000). Heterogeneous preferences regarding global climate change. *Review of Economics and Statistics 82*(4), 616–624.

Lochner, L. (2007). Individual perceptions of the criminal justice system. *American Economic Review 97*(1), 444–460.

Low, C. (2022). Pricing the biological clock: The marriage market costs of aging to women. Journal of Labour Economics (forthcoming).

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics 3*(3), 205–228.

Manski, C. F. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of Econometrics 27*(3), 313–333.

Manski, C. F. (1999). Analysis of choice expectations in incomplete scenarios. *Journal of Risk and Uncertainty 19*(1-3), 49–66.

Manski, C. F. (2004). Measuring expectations. *Econometrica 72*(5), 1329–1376.

Manski, C. F. and F. Molinari (2010). Rounding probabilistic expectations in surveys. *Journal of Business and Economic Statistics 28*(2), 219–231. PMID: 20368764.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*, New York, pp. 105–142. Academic Press.

McFadden, D. and K. Train (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics 15*(5), 447–470.

McKenzie, D., J. Gibson, and S. Stillman (2013). A land of milk and honey with streets paved with gold: Do emigrants have over-optimistic expectations about incomes abroad? *Journal of Development Economics 102*, 116–127. Migration and Development.

Miller, G., de Paula, and C. Valente (2020). Subjective expectations and demand for contraception. Working Paper 27271, National Bureau of Economic Research.

Morita, T. and S. Managi (2015). Consumers' willingness to pay for electricity after the great east japan earthquake. *Economic Analysis and Policy 48*, 82–105. Energy.

Papke, L. E. and J. M. Wooldridge (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics 11*(6), 619–632.

Pedersen, L. B., M. R. Mørkbak, and R. Scarpa (2020). Handling resolvable uncertainty from incomplete scenarios in future doctors' job choice – probabilities vs discrete choices. *Journal of Choice Modelling 34*, 100199.

Revelt, D. and K. Train (1998). Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of Economics and Statistics 80*(4), 647–657.

Ruder, A. I. and M. Van Noy (2017). Knowledge of earnings risk and major choice: Evidence from an information experiment. *Economics of Education Review 57*, 80–90.

Ryan, M. (1999). Using conjoint analysis to take account of patient preferences and go beyond health outcomes: an application to in vitro fertilisation. *Social Science and Medicine 48*(4), 535–546.

Scarpa, R., C. Bazzani, D. Begalli, and R. Capitello (2021). Resolvable and near-epistemic uncertainty in stated preference for olive oil: An empirical exploration. *Journal of Agricultural Economics 72*(2), 335–369.

Shoyama, K., S. Managi, and Y. Yamagata (2013). Public preferences for biodiversity conservation and climate-change mitigation: A choice experiment using ecosystem services indicators. *Land Use Policy 34*, 282–293.

Small, K. A., C. Winston, and J. Yan (2005). Uncovering the distribution of motorists' preferences for travel time and reliability. *Econometrica 73*(4), 1367–1382.

Soekhai, V., E. W. de Bekker-Grob, A. R. Ellis, and C. M. Vass (2019). Discrete choice experiments in health economics: past, present and future. *Pharmacoeconomics 37*, 201–226.

Summerfield, M., B. Garrard, M. Nesa, R. Kamath, N. Macalalad, N. Watson, R. Wilkins, and M. Wooden (2022). HILDA User Manual – Release 22. Melbourne Institute: Applied Economic and Social Research, University of Melbourne.

Tarozzi, A., A. Mahajan, B. Blackburn, D. Kopf, L. Krishnan, and J. Yoong (2014). Micro-loans, insecticide-treated bednets, and malaria: Evidence from a randomized controlled trial in orissa, india. *American Economic Review 104*(7), 1909–41.

Train, K. and M. Weeks (2005). Discrete choice models in preference space and willingness-to-pay space. In R. Scarpa and A. Alberini (Eds.), *Applications of Simulation Methods in Environmental and Resource Economics*, Dordrecht, pp. 1–16. Springer Netherlands.

Train, K. E. (2009). *Discrete Choice Methods with Simulation* (2nd ed.). Cambridge University Press.

Walker, J. L., M. Ben-Akiva, and D. Bolduc (2007). Identification of parameters in normal error component logit-mixture (neclm) models. *Journal of Applied Econometrics 22*(6), 1095–1125.

Wiswall, M. and B. Zafar (2015). Determinants of College Major Choice: Identification using an Information Experiment. *Review of Economic Studies 82*(2), 791–824.

Wiswall, M. and B. Zafar (2018). Preference for the Workplace, Investment in Human Capital, and Gender. *Quarterly Journal of Economics 133*(1), 457–507.

Yan, J. and H. I. Yoo (2019). Semiparametric estimation of the random utility model with rank-ordered choice data. *Journal of Econometrics 211*(2), 414–438.

Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources 48*(3), 545–595.

Online Appendix

Alternative Models of Preference Heterogeneity for Elicited Choice Probabilities

Nathan Kettlewell
Economics Department, University of Technology Sydney, Australia
Matthew J. Walker
Economics Subject Group, Newcastle University, UK
and
Hong Il Yoo
Loughborough Business School, Loughborough University, UK

Appendix A: Likelihood Functions

Continuing with our notation in Section 2 of the main text, let V_{njt} denote the utility index for alternative $j \in \{1, 2\}$ that respondent $n \in \{1, 2, \cdots, N\}$ evaluates in choice scenario $t \in \{1, 2, \cdots, T\}$. This index is a function of preference parameters $\boldsymbol{\theta}_n$ which are randomly distributed between individuals. The exact content of $\boldsymbol{\theta}_n$ varies from model to model, for example $\boldsymbol{\theta}_n = \{\boldsymbol{\beta}_n, \lambda_n\}$ for the NMXL model in equation (3) and $\boldsymbol{\theta}_n = \{\boldsymbol{\beta}_n, \lambda_n, \sigma_n\}$ for the GMNL-II model in equation (5).

Consider first respondents who completed the standard choice tasks. Conditional on $\boldsymbol{\theta}_n$, the likelihood of respondent n's choice in task t is given by a logit density function of the form

$$\mathrm{m}_{nt}[d_{n2t}; \boldsymbol{\theta}_n] = \Big(\frac{\exp[V_{n2t}]}{(\exp[V_{n1t}] + \exp[V_{n2t}])}\Big)^{d_{n2t}} \Big(\frac{\exp[V_{n1t}]}{(\exp[V_{n1t}] + \exp[V_{n2t}])}\Big)^{(1-d_{n2t})} \tag{A1}$$

where d_{n2t} is a binary indicator which is equal to 1 if their choice is alternative 2 and 0 if alternative 1. Let $\mathrm{f}[\boldsymbol{\theta}_n; \boldsymbol{\Theta}]$ denote a density function that describes the population distribution of $\boldsymbol{\theta}_n$ as a function of population-level distributional parameters $\boldsymbol{\Omega}$, such as the population mean and standard deviation in the case of a normally distributed random parameter. The unconditional likelihood of T observations on respondent n is then given by

$$\mathrm{M}_n[\boldsymbol{\Theta}] = \int \prod_{t=1}^{T} \mathrm{m}_{nt}[d_{n2t}; \boldsymbol{\theta}_n]\mathrm{f}[\boldsymbol{\theta}_n; \boldsymbol{\Theta}]d\boldsymbol{\theta}_n \tag{A2}$$

which does not have an analytic expression. We construct a simulated analogue to the unconditional sample log-likelihood function, $\sum_{n=1}^{N} \ln\big[\mathrm{M}_n[\boldsymbol{\Theta}]\big]$, by following the procedure outlined in Train (2009, §6.7). Our MSL estimates of $\boldsymbol{\Theta}$ are computed by maximizing this simulated analogue.

Consider next respondents who completed the probability elicitation task. The conditional likelihood of respondent n's stated choice probability in task t is given by the beta density function $\mathrm{l}_{nt}[y_{n2t}; \boldsymbol{\theta}_n, \phi]$ in equation (9) of the main text. Accordingly, the unconditional likelihood of T observations on respondent n takes the form of

$$\mathrm{L}_n[\boldsymbol{\Theta}, \phi] = \int \prod_{t=1}^{T} \mathrm{l}_{nt}[y_{n2t}; \boldsymbol{\theta}_n, \phi]\mathrm{f}[\boldsymbol{\theta}_n; \boldsymbol{\Theta}]d\boldsymbol{\theta}_n \tag{A3}$$

which does not have an analytic expression either. We apply the same procedure as we do with the choice tasks to simulate the sample log-likelihood function, $\sum_{n=1}^{N} \ln\big[\mathrm{L}_n[\boldsymbol{\Theta}, \phi]\big]$, and compute the MSL estimates of $\boldsymbol{\Theta}$ and ϕ.

Appendix B: Participant Instructions

B1. Instructions for Standard Choice Tasks

For the next set of questions we would like you to imagine that you are purchasing a new car.

You will be shown 8 imaginary situations. In each situation, you will have a choice between 2 different cars. Each car is described with a list of 8 features. You are asked to look at all the features and think about how important they are to you. You task is to indicate which car you would purchase if you had to choose between these 2 options.

Some of the car features may be unfamiliar to you. This information sheet explains what each feature means. You may keep it open as you complete the task or you may choose to print it.

There are no right or wrong answers; we are simply interested in your views.

Scenario 1 of 8: Please select which car you would purchase.

	Car A	Car B
Purchase price	$40,000	$50,000
Fuel type	Hydrogen Fuel Cell	Battery Electric
Fuel cost per 100km	$13	$6.50
Car size	Large	Small
CO2 emmissions (% average car)	50%	0%
Driving range	400 km	700 km
Fuel availability (% of service stations)	20%	100%
Battery recharging time	Not applicable	10 minutes

Car A ○ Car B ○

Figure B1: Choice set example

B2. Instructions for Probability Elicitation Tasks

For the next set of questions we would like you to imagine that you are purchasing a new car.

You will be shown 8 imaginary situations. In each situation, you will have a choice between 2 different cars. Each car is described with a list of 8 features. You are asked to look at all the features and think about how important they are to you.

For each scenario, you are asked to indicate the chance in percentage terms of choosing each of the 2 options. The chance of choosing each option should be a number between 0 and 100 and the chances given to the two alternatives should add up to 100. For example, if you give a 5% chance to one alternative it means that there is almost no possibility that you will choose that option. On the other hand, if you give an 80% or over chance to an option it means that you will almost surely choose it.

Some of the car features may be unfamiliar to you. This information sheet explains what each feature means. You may keep it open as you complete the task or you may choose to print it.

There are no right or wrong answers; we are simply interested in your views.

Scenario 1 of 8: Please indicate how likely you are to pick each option as a percentage from 0 to 100. Percentages must add up to 100.

	Car A	**Car B**
Purchase price	$40,000	$50,000
Fuel type	Hydrogen Fuel Cell	Battery Electric
Fuel cost per 100km	$13	$6.50
Car size	Large	Small
CO2 emmissions (% average car)	50%	0%
Driving range	400 km	700 km
Fuel availability (% of service stations)	20%	100%
Battery recharging time	Not applicable	10 minutes

Car A	0
Car B	0
Total	0

Figure B2: Choice set example

B3. Information sheet for both types of tasks

In both sets of instructions, the word 'information sheet' is underlined. Clicking it will open a PDF document with the following information.

Car sizes

Small car – Engine size typically 1.3-2.0 cubic litres and length less than 7.5 meters. Popular models of small car include Toyota Corolla and Mazda 3.

Medium car – Engine size typically 2.0-3.0 cubic litres and length between 7.5 and 8.3 meters. Popular examples of medium sized cars include Mazda 6 and Subaru Liberty.

Large car – Engine size typically >3.0 cubic litres and length greater than 8.3 meters. Popular example of large cars includes Toyota Camry and Holden Commodore. For the purpose of this study, sports utility vehicles and 4WDs can be treated as large cars.

Fuel types

Conventional fuel – In this study, conventional fuel means either petrol (unleaded) or diesel. These vehicles used standard internal combustion engines.

Liquid petroleum gas (LPG) – These vehicles use the same engine as conventional fuel vehicles with small modifications to the fuel system to accept liquified petroleum gas. For the purpose of this study, assume that LPG vehicles are dedicated LPG (i.e. not dual fuel).

Hybrid electric vehicles – These vehicles run on a combination of electricity and conventional fuel. They do not require plugging-in to recharge and switch between an internal combustion engine and electric propulsion system.

Plug-in hybrid electric vehicles – Similar to hybrid electric vehicles, plug-in hybrid electric vehicles run on a combination of electricity and conventional fuel. However, they can be plugged into power outlets to recharge, which allows the electricity store to be topped up while parked reducing reliance on the internal combustion system.

Battery electric vehicles – These vehicles run on electricity stored in rechargeable battery packs and must be recharged at a power outlet. They use an electric motor (not use an internal combustion engine).

Hydrogen fuel cell electric vehicles – These vehicles use an electric motor powered by a hydrogen fuel cell (a tank of compressed hydrogen). The tanks must be refuelled from a hydrogen refuelling station.

Driving range

This is the maximum distance the car can be driven before refuelling/recharging.

Appendix C: Additional Empirical Results

Table C1: Standard deviations of residual taste heterogeneity for population-level preference estimates using choice probabilities (paired with Table 3).

	F-NMXL	F-GMNL-II	F-GMNL
Medium size	0.029	0.302***	0.323***
	(0.140)	(0.076)	(0.055)
Large size	0.209***	0.112***	0.110***
	(0.070)	(0.039)	(0.027)
Liquid Petroleum Gas	0.197	0.636***	0.546***
	(0.133)	(0.094)	(0.083)
Hybrid Electric	0.260	0.099*	0.178***
	(0.221)	(0.060)	(0.048)
Hydrogen Fuel Cell	0.257	0.075	0.009
	(0.165)	(0.080)	(0.117)
Plug-in Hybrid Electric	0.311**	0.279***	0.374***
	(0.123)	(0.057)	(0.056)
Battery Electric	0.521***	0.730***	0.619***
	(0.194)	(0.084)	(0.095)
Fuel cost	0.039***	0.053***	0.049***
	(0.007)	(0.006)	(0.006)
CO2 Emissions	0.298*	0.470***	0.368***
	(0.176)	(0.106)	(0.053)
Recharge time	0.058**	0.143***	0.162***
	(0.024)	(0.021)	(0.020)
Refill availability	0.309**	0.135	0.169
	(0.152)	(0.171)	(0.114)
Driving range	0.057***	0.095***	0.080***
	(0.017)	(0.010)	(0.010)
-1 * Price	0.186***	0.216***	0.191***
	(0.018)	(0.020)	(0.019)

Notes: Standard errors in parentheses are clustered at the respondent level (202 respondents).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C2: Population-level estimates in WTP space using choice probabilities.

	F-WTP	F-NWTP	
		Mean	SD
Medium size	0.522	0.334	0.945***
	(0.357)	(0.308)	(0.295)
Large size	0.852**	0.305	0.857**
	(0.341)	(0.391)	(0.412)
Liquid Petroleum Gas	-0.551	-0.973**	1.011**
	(0.518)	(0.399)	(0.409)
Hybrid Electric	0.833	-0.089	0.816
	(0.570)	(0.696)	(0.712)
Hydrogen Fuel Cell	0.429	0.076	1.808***
	(0.540)	(0.567)	(0.373)
Plug-in Hybrid Electric	-0.489	-1.211**	1.749***
	(0.694)	(0.599)	(0.407)
Battery Electric	-0.131	-0.519	2.772***
	(0.686)	(0.512)	(0.639)
Fuel cost	-0.179***	-0.175***	0.178***
	(0.036)	(0.029)	(0.033)
CO2 Emissions	-1.418***	-0.492*	3.254***
	(0.435)	(0.256)	(0.544)
Recharge time	-0.208**	-0.284***	0.289***
	(0.089)	(0.084)	(0.080)
Refill availability	1.682***	1.960***	2.480***
	(0.481)	(0.376)	(0.306)
Driving range	0.235***	0.171***	0.326***
	(0.076)	(0.043)	(0.057)
-1 * Price			0.962***
			(0.076)
ϕ	1.249***	2.372***	
	(0.107)	(0.317)	
σ	0.145***	0.307***	0.379***
	(0.021)	(0.036)	(0.064)
No. of parameters	14	27	
No. of choice sets	1612	1612	
LL	538.296	682.318	
AIC	-1048.592	-1310.637	
BIC	-973.199	-1165.236	

Notes: Standard errors in parentheses are clustered at the respondent level (202 respondents).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C3: Population-level preference estimates for AFVs using stated choices.

	MNL	NMXL	GMNL-II[1]	GMNL[2]
Medium size	0.334***	0.417***	1.505	0.399***
	(0.058)	(0.092)	(1.068)	(0.102)
Large size	0.338***	0.378***	1.419	0.396***
	(0.063)	(0.100)	(1.070)	(0.113)
Liquid Petroleum Gas	-0.015	-0.234	-0.947	-0.236
	(0.101)	(0.187)	(0.808)	(0.212)
Hybrid Electric	0.059	-0.024	0.247	0.104
	(0.102)	(0.170)	(0.467)	(0.207)
Hydrogen Fuel Cell	-0.056	-0.131	-0.219	-0.119
	(0.099)	(0.170)	(0.507)	(0.174)
Plug-in Hybrid Electric	-0.164	-0.428**	-1.129	-0.349*
	(0.101)	(0.175)	(0.858)	(0.191)
Battery Electric	-0.158*	-0.414**	-1.433	-0.390**
	(0.090)	(0.165)	(1.013)	(0.168)
Fuel cost	-0.050***	-0.090***	-0.382	-0.100***
	(0.004)	(0.010)	(0.274)	(0.021)
CO2 Emissions	-0.300***	-0.516***	-2.636	-0.612***
	(0.069)	(0.118)	(1.703)	(0.188)
Recharge time	-0.096***	-0.150***	-0.735	-0.184***
	(0.015)	(0.027)	(0.595)	(0.053)
Refill availability	0.661***	1.037***	3.794	1.057***
	(0.075)	(0.138)	(2.661)	(0.192)
Driving range	0.091***	0.143***	0.566	0.156***
	(0.009)	(0.018)	(0.401)	(0.028)
-1 * Price	0.265***	0.532***	2.102	0.533***
	(0.018)	(0.051)	(1.666)	(0.098)
τ			-1.688***	-0.682**
			(0.467)	(0.282)
γ				0.662*
				(0.365)
No. of parameters	13	26	20	28
No. of choice sets	4200	4200	4200	4200
LL	-2355.1	-2232.7	-2218.2	-2227.1
AIC	4736.2	4517.5	4476.5	4510.1
BIC	4818.7	4682.4	4603.3	4687.7

Notes: Standard errors in parentheses are clustered at the respondent level (525 respondents).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
[1]The variance parameters on the seven dummy variables are constrained to zero.
[2]Simulation of the likelihood function is based on only 50 Halton draws, instead of 500 draws used elsewhere.

Table C4: Standard deviations of residual taste heterogeneity for population-level preference estimates using stated choices (paired with Table C3).

	(1) NMXL	(2) GMNL-II[1]	(3) GMNL[2]
Medium size	0.063		0.026
	(0.101)		(0.114)
Large size	0.489***		0.197
	(0.179)		(0.215)
Liquid Petroleum Gas	1.113***		0.699
	(0.416)		(0.598)
Hybrid Electric	0.571		0.579
	(0.545)		(0.423)
Hydrogen Fuel Cell	0.011		0.110
	(0.234)		(0.252)
Plug-in Hybrid Electric	0.278		0.316*
	(0.361)		(0.192)
Battery Electric	0.372*		0.302
	(0.193)		(0.195)
Fuel cost	0.098***	0.383	0.077***
	(0.014)	(0.296)	(0.013)
CO2 Emissions	1.316***	5.209	1.327***
	(0.226)	(3.956)	(0.268)
Recharge time	0.225***	1.058	0.231***
	(0.045)	(0.850)	(0.039)
Refill availability	1.276***	4.438	0.763*
	(0.227)	(3.355)	(0.439)
Driving range	0.143***	0.533	0.115***
	(0.024)	(0.384)	(0.040)
-1 * Price	0.464***	1.917	0.414***
	(0.049)	(1.522)	(0.050)

Notes: Standard errors in parentheses are clustered at the respondent level (525 respondents).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
[1]The variance parameters on the seven dummy variables are constrained to zero.
[2]Simulation of the likelihood function is based on only 50 Halton draws, instead of 500 draws used elsewhere.

Table C5: Population-level estimates in WTP space using stated choices.

	WTP	NWTP[1]	
		Mean	SD
Medium size	1.263***	0.634***	
	(0.226)	(0.161)	
Large size	1.278***	0.530**	
	(0.222)	(0.252)	
Liquid Petroleum Gas	-0.058	-1.274***	
	(0.380)	(0.341)	
Hybrid Electric	0.221	-0.268	
	(0.385)	(0.341)	
Hydrogen Fuel Cell	-0.213	-0.407	
	(0.375)	(0.275)	
Plug-in Hybrid Electric	-0.621	-1.065***	
	(0.383)	(0.348)	
Battery Electric	-0.598*	-1.144***	
	(0.346)	(0.353)	
Fuel cost	-0.189***	-0.203***	0.138***
	(0.019)	(0.015)	(0.015)
CO2 Emissions	-1.133***	-0.990***	1.860***
	(0.266)	(0.207)	(0.189)
Recharge time	-0.362***	-0.374***	0.419***
	(0.058)	(0.062)	(0.072)
Refill availability	2.497***	1.659***	0.684***
	(0.319)	(0.368)	(0.160)
Driving range	0.344***	0.287***	0.244***
	(0.039)	(0.031)	(0.028)
-1 * Price			1.990***
			(0.424)
σ	0.265***	12.11	86.931
	(0.018)	(17.50)	(200.196)
No. of parameters	13	19	
No. of choice sets	420	420	
LL	-2355.122	-2258.031	
AIC	4736.244	4554.062	
BIC	4818.700	4674.576	

Notes: Standard errors in parentheses are clustered at the respondent level (525 respondents).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
[1]The variance parameters on the seven dummy variables are constrained to zero.

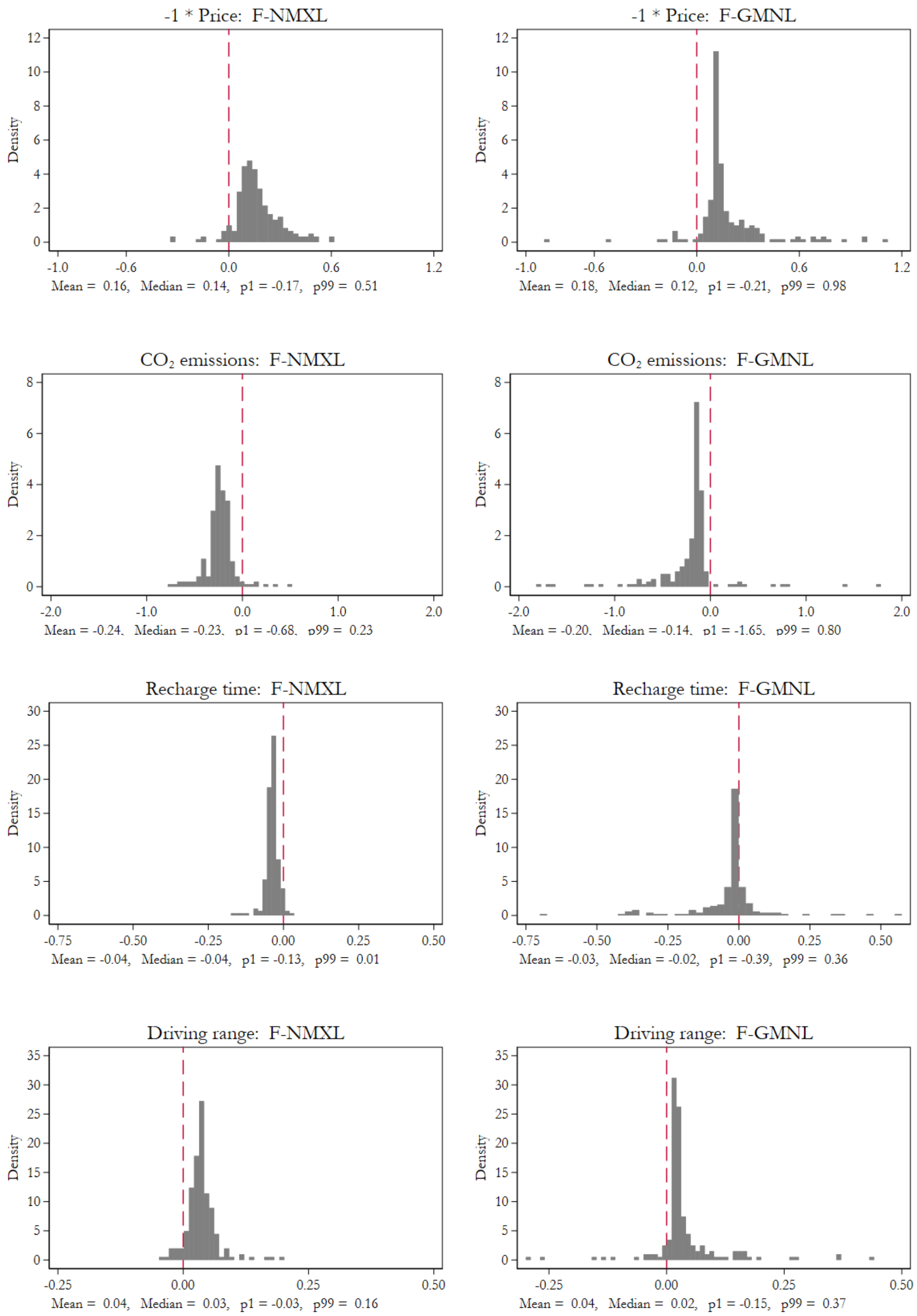

Figure C1: Posterior distributions of individual-level preference parameters for the choice attributes *Price*, *Recharge time*, *Driving range* and *CO_2 emissions*) in F-NMXL (column 1) and F-GMNL (column 2). Paired with Figure 2.

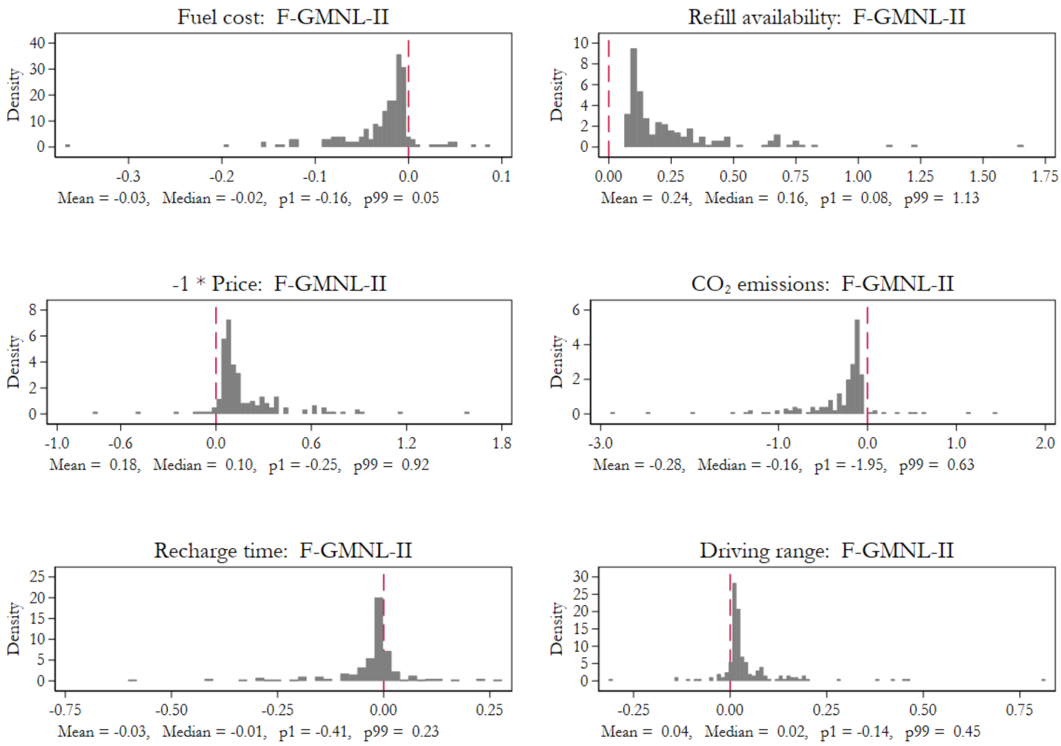

Figure C2: Posterior distributions of individual-level preference parameters for the numeric choice attributes in F-GMNL-II. Paired with Figure 2.