Received: 13 March 2021 | Accepted: 3 December 2022

DOI: 10.1111/ajae.12367

ARTICLE

# A new empirical approach for mitigating exploding implicit prices in mixed multinomial logit models

**Romain Crastes dit Sourd**

Centre for Decision Research, Management Department, Leeds University Business School, University of Leeds, Leeds, UK

**Correspondence**
Romain Crastes dit Sourd, Centre for Decision Research, Management Department, Leeds University Business School, University of Leeds, Leeds, UK.
Email: r.crastesditsourd@leeds.ac.uk

**Abstract**

This paper introduces a new shifted negative log-normal distribution for the price parameter in mixed multinomial logit models. The new distribution, labeled as the μ-shifted negative log-normal distribution, has desirable properties for welfare analysis and in particular a point mass that is further away from zero than the negative log-normal distribution. This contributes to mitigating the "exploding" implicit prices issue commonly found when the price parameter is specified as negative log-normal and the model is in preference space. The new distribution is tested on five stated preference datasets. Comparisons are made with standard alternative approaches such as the willingness-to-pay (WTP) space approach. It is found that the μ-shifted distribution yields substantially lower mean marginal WTP estimates compared to the negative log-normal specification and similar to the values derived from models estimated in WTP-space with flexible distributions, while at the same time fitting the data as well as the negative log-normal specification.

**KEYWORDS**

choice modeling, mixed logit, non-market valuation, random utility, utility in WTP-space

**JEL CLASSIFICATION**

Q51, C35

## 1 | INTRODUCTION

Choice models fitted to stated and revealed preference (SP and RP) data for non-market valuation commonly employ random utility choice models with continuous random heterogeneity in preferences across decision makers (Alfnes et al., 2006; Du et al., 2017; Peterson et al., 2015; Uz et al., 2022). This paper contributes to a series of methods developed to derive reasonable distributions of marginal willingness to pay (mWTP) from such models, commonly known as mixed logit models. Whether such distributions are reasonable generally depends on how heavy tailed they are (Mariel, Demel, & Longo, 2021; Scarpa, Thiene, & Marangon, 2008; Sonnier et al., 2007; Train & Weeks, 2005). Heavy-tailed distributions often arise when the numerator of the ratio of distributions used to simulate mWTP can take values very proximate to zero, leading to exploding implicit prices' (Giergiczny et al., 2012).

In most cases, assuming no random heterogeneity in preferences across decision makers for the price attribute yields reasonable willingness-to-pay (WTP) values, but such an assumption is behaviorally implausible and might be rejected against competing specifications. Common alternatives include either using a log-normal distribution (2LN) for the price attribute, which puts the analyst at risk of an exploding ratio problem or estimating the model in WTP space (Train & Weeks, 2005). Models estimated in WTP space (as opposed to preference or utility space) use the monetary attribute to rescale the formulation of utility (Thiene & Scarpa, 2009). This allows us to directly specify the WTP distribution for each of the (non-monetary) attributes instead of deriving it through a ratio of distributions. This property helps to avoid exploding implicit prices. It can also allow better testing of how specific restrictions on the variance–covariance of the coefficient distributions affect model fit (Rungie et al., 2014; Thiene & Scarpa, 2009). Models in preference space have often been found to fit the data better than models in WTP space, because the distributional assumptions allowed by working in preference space are simply more likely to better accommodate extreme preferences for a given attribute (or lower price sensitivity for some decision makers). This leads Train and Weeks (2005) to suggest that "*research is needed to identify distributions that fit the data better when applied in WTP-space and/or provide more reasonable distributions of WTP when applied in preference space*."

Several authors have engaged with this proposition in the past. Most efforts have consisted in finding distributions that yield more reasonable mWTP estimates in preference space. Svenningsen and Jacobsen (2018) have used a normal distribution for the monetary attribute. However, Daly et al. (2012) demonstrated that this can prevent the existence of moments for the mWTP distribution of the non-monetary attributes. A different approach has been suggested by Giergiczny et al. (2012), who introduced a cost–income variable in an attempt to shift the (negative log-normal) distribution of the cost away from zero. However, this does not guarantee that the additional interaction variable behaves as intended, in addition to introducing more data requirements. The lack of an existing solution for systematically preventing exploding implicit prices when estimating mixed logit models in preference space with a randomly distributed monetary parameter can severely complicate model selection. As Mariel, Hoyos, et al. (2021) put it, "*choosing a model is ultimately based on the researcher's own judgement, which is informed by several, sometimes contrasting criteria and the purpose of the research*."

In this paper, we propose a new approach for deriving more reasonable WTP distributions from models estimated in preference space. It is based around the shifted log-normal distribution, which is also known as the 3LN (Kalecki, 1945; Sangal & Biswas, 1970). The shifted log-normal distribution is simply obtained by adding a shift parameter to a classic log-normal distribution, as defined in the remainder of this paper. However, such a specification is unable to provide moments for the mWTP distributions when the shift parameter is found to be positive at model convergence. In addition, a recent contribution from McFadden (2022) shows that the three-parameter log-normal also suffers from identification issues. Our proposition consists of constraining the shift parameter to be equal to the exponential of the mean of the underlying normal distribution. We label this distribution the μ-shifted distribution.

We evaluate the performances of the μ-shifted distribution on five datasets from the non-market valuation and transportation literature. For each dataset, we estimate a series of models and compare new and existing common parameterizations. We use the recent guidelines for comparing models authored by Mariel, Hoyos, et al. (2021) as a framework for comparing models and WTP estimates. We find in particular that the new μ-shifted parameterization consistently leads to goodness-of-fit measures, which are nearly identical or equal to the model exhibiting the best fit for a given dataset. At the same time, the proposed approach yields WTP estimates that are not significantly different from those derived from models estimated in WTP space with flexible mixture distributions and far more reasonable than the values derived from preference-space models featuring a log-normally distributed price parameter.

This paper is organized as follows. The next section presents the new shifted distribution. Section 3 introduces the data used in this analysis. Section 4 describes the framework for empirical testing used to compare the new shifted distribution to existing alternatives in the literature. Section 5 presents and discusses results. Section 6 discusses other features of the new shifted distribution for analysts and concludes.

## 2 | METHODS

### 2.1 | The mixed multinomial logit framework

We start by describing the well-known mixed multinomial logit (MMNL) specification in preference space. Let $U_{int}$ be the utility that respondent $n$ derives from alternative $i$ in choice situation $t$. The utility includes a set of modeled components and a random component $\varepsilon_{int}$, which follows a Type 1 extreme value distribution. We have:

$$U_{int} = ASC_i + \beta_n' \boldsymbol{x}_{int} + \rho_n price_{int} + \varepsilon_{int} \tag{1}$$

where $\beta_n$ is a vector of taste coefficients (excluding the sensitivity for the price), $\boldsymbol{x}_{int}$ a vector of attributes for alternative $i$, $\rho_n$ captures price sensitivity for respondent $n$, and $price_{int}$ corresponds to the value of the price attribute for alternative $i$ faced by respondent $n$ in choice situation $t$. We include alternative specific constants (ASCs) for all but one of the alternatives. The probability that respondent $n$ chooses a given alternative $i$ conditional on $\beta_n$, $\rho_n$ and the ASCs in choice situation $t$ corresponds to the well know multinomial logit (MNL) probability. The elements in $\beta_n$ and $\rho_n$ can be allowed to vary randomly across respondents. A common assumption in non-market valuation is to assume that the elements in $\beta_n$ are normally distributed:

$$\beta_{kn} = \mu_k + \sigma_k \zeta_{kn} \tag{2}$$

where $\mu_k$ corresponds to the mean and the standard deviation of the random parameter. In this example, $\zeta_{kn}$ is a vector of standard normal draws $N(0, 1)$ for respondent $n$. This is a very common assumption for non-monetary parameters in non-market valuation, but other distributions (Uniform, Triangular, etc.) can be used too. In some cases, the distribution of a given attribute needs to be constrained for behavioral reasons or for ensuring the existence of moments for the distribution of mWTP estimates. This is, for example, the case for the price attribute, which is commonly assumed to be log-normal (Mariel, Demel, & Longo, 2021):

$$\rho_n = -e^{\left(\mu_{price} + \sigma_{price} \zeta_{price,n}\right)} \tag{3}$$

where $\zeta_{price,n}$ is again a set of standard normal draws for respondent $n$. This assumption can have undesirable features when it comes to deriving mWTP distributions. Indeed, mWTP distributions are derived for each non-monetary attribute by calculating the following ratio using a large number of draws (in this paper, all mWTP distributions are simulated using 10 million draws):

$$mWTP_{kn} = -\frac{\beta_{kn}}{\rho_n} \tag{4}$$

The log-normal distribution (2LN) has a point-mass near zero, which means that the denominator of Equation (4) is likely to reach extremely small values, leading to very large mWTP estimates. This is particularly prone to happen when $\sigma_{price}$ increases, that is when there is a large amount of heterogeneity in preferences for the monetary attribute in the survey sample. These large values have been found to have a considerable impact on mean mWTP estimates. This problem is known as the "exploding implicit price" problem in the literature (see Giergiczny et al., 2012) and the mean mWTP values derived from such models have been branded as unreasonable or counterintuitive based on expert knowledge (Scarpa, Thiene, & Train, 2008). A series of alternative parameterizations have been suggested in the literature to circumvent this issue. However, each proposition has its drawbacks and a universal solution does not exist. The right modeling approach is always case specific and depends on the data at hand.

## 2.2 | Alternative parameterizations

### WTP space

A popular alternative to the parameterization presented above is the WTP-space approach. The WTP-space approach was first suggested by Train and Weeks (2005), although the concept was put forward in Cameron (1988) in the context of referendum contingent valuation data. A model parameterized in WTP-space is obtained by reformulating Equation (1) as follows:

$$U_{int} = \rho_n \cdot \left(ASC_i + \beta_n' \boldsymbol{x}_{int} - price_{int}\right) + \varepsilon_{int} \tag{5}$$

The elements of $\beta_n'$ are now directly interpretable as WTP estimates. In other words, models specified in WTP space make it possible to directly specify the mWTP distribution for each attribute, including flexible mixture distributions. In particular, following Fosgerau and Mabit (2013)[1] for models estimated using simulated maximum likelihood, it is possible to specify the mWTP distribution for attribute $k$ (labeled as $\beta_{kn}$) as a mixture distribution using a second-order polynomial in a standard normal distribution:

$$\beta_{kn} = \mu_k + \sigma 1_k \zeta_{kn} + \sigma 2_k (\zeta_{kn})^2 \tag{6}$$

where $\mu_k$, $\sigma 1_{k,}$ and $\sigma 2_k$ are parameters to be estimated and $\zeta_{kn}$ is a set of standard normal draws, whereas is the square of the same set of draws. In this context, $\sigma 1_k$ and $\sigma 2_k$ are shape parameters of the mWTP distribution for attribute $k$. The mean mWTP is obtained by simulating $\beta_{kn}$ over a large number of draws and computing the average of the resulting distribution.

Directly specifying the mWTP distribution for each attribute instead of simulating ratios of distributions helps to avoid issues with heavy-tailed mWTP distributions as previously stated. Moreover, models specified in WTP space can allow for the direct and efficient testing of hypotheses on mWTP distributions (under certain distributional assumptions). Applications include testing the positioning of quantiles for pricing strategies (Thiene & Scarpa, 2009) or imposing specific restrictions on the correlation structure of the random parameters as suggested by Rungie et al. (2011) (see also Rungie et al. (2014)).

[1]See also Scarpa et al. (2008), Train (2016) and Scarpa et al. (2021) for alternative ways of introducing flexible taste distributions.

On the other hand, models estimated in WTP space can only provide more reasonable mWTP estimates for the numeraire used for normalizing the model, although there are contexts where decision makers face choices defined by two or more cost attributes (or other numeraires, see for example Meginnis et al., 2020). Welfare measures for the other numeraires (not used for normalizing the model) need to be derived by means of simulated ratios of distributions, in a similar fashion to what is done in preference space, and in which case the point of estimating the model in WTP space is partially lost as this introduces a risk of "exploding ratio."

### Other parameterizations in preference space

Other solutions often found in the literature consist in fixing $\sigma_{price}$ to zero (Rigby et al., 2009) or using other distributions for $\rho$ such as the negative log-uniform distribution, where $\zeta_{price,n}$ in Equation (3) is assumed to be distributed $U(0, 1)$ instead (Czine et al., 2020). In this paper, we compare the aforementioned parameterizations to a new distribution for the price attribute based on the shifted log-normal distribution as previously introduced.

## 2.3 | Shifted negative log-normal distributions

### The three-parameter log-normal distribution (3LN)

The original shifted negative log-normal distribution described by Sangal and Biswas (1970) is obtained by adding one shift parameter to the 2LN distribution described in Equation (3):

$$\rho_n = \kappa - e^{\left(\mu_{price} + \sigma_{price}\zeta_{price,n}\right)} \tag{7}$$

where $\kappa$ is a shift parameter to be estimated. In the remainder of this paper, we refer to this specification as the 3LN. Introducing $\kappa$ is expected to contribute to shift the denominator in Equation (4) away from zero and hence mitigate the "exploding ratio" issue. However, this can only be the case if $\kappa$ is found to be negative. In the event where a positive $\kappa$ is found, the model cannot be used for welfare analysis as the distribution of $\rho$ can span both sides of zero. This is a problem because this would suggest that no finite mWTP moments exist (this issue is thoroughly documented in Daly et al., 2012). Constraining $\kappa$ to be negative via an exponentiation is also undesirable. More precisely, if $\kappa$ is unconstrained and found to be positive, its constrained counterpart is likely to lead to a shift that is extremely close to zero, which in turn can lead to issues during model estimation, in addition to having no or very little impact when it comes to moving the denominator of Equation (4) away from zero.

Another concern regarding the use of a 3LN distribution for the price attribute, which has been recently pointed out by McFadden (2022) is that the parameters $\mu$ and $\kappa$ described in Equation (7) are collinear and were found to be "*poorly estimated*" in a series of Monte-Carlo simulations. The authors state that "*this is unsurprising since both act to determine the location of the* [...] *distribution, and have effects that are separately identified only through the behavior of consumers facing the most extreme high prices*." In this paper, we propose a new specification for the 3LN distribution, which seeks to mitigate these issues.

### The $\mu$-shifted log-normal distribution

Our proposition consists in replacing the shift parameter $\kappa$ of the 3LN distribution by the exponential of the mean of the underlying normal distribution, that is $-e^{\left(\mu_{price}\right)}$. Such a specification has the

TABLE 1 Negative 2LN distribution versus $\mu$LN

| Measure | 2LN | $\mu$LN |
|---|---|---|
| Mean | $-e^{\mu+\sigma^2/2}$ | $-e^{\mu}-e^{\mu+\sigma^2/2}$ |
| Variance | $\left(e^{\sigma^2}-1\right)\cdot e^{2\mu+\sigma^2}$ | $\left(e^{\sigma^2}-1\right)\cdot e^{2\mu+\sigma^2}$ |
| Mode | $-e^{(\mu-\sigma^2)}$ | $-e^{\mu}-e^{(\mu-\sigma^2)}$ |
| P.D.F. | $-\frac{1}{p\sigma\sqrt{2\pi}}\cdot e^{-\frac{(log(-p)-\mu)^2}{2\sigma^2}}$ | $-\frac{1}{(p+e^{\mu})\sigma\sqrt{2\pi}}\cdot e^{-\frac{(log(-p-e^{\mu})-\mu)^2}{2\sigma^2}}$ |
| Support | $]-\infty,0[$ | $]-\infty,-e^{\mu}[$ |

*Note*: comparison of the 2LN and $\mu$LN distributions in terms of mean, variance, mode, probability density function, and support.
Abbreviation: PDF, Probability Density Function.

desirable properties of only relying on μ to determine the location of the distribution while anchoring the shift to the expected value of the logarithm of the sensitivity for the price attribute, which ensures that the shift does not become positive or very close to zero. We label the resulting distribution the μ-shifted log-normal distribution ($\mu$LN). Formally, we propose:

$$\rho_n = -e^{\left(\mu_{price}\right)} - e^{\left(\mu_{price}+\sigma_{price}\zeta_{price,n}\right)} \tag{8}$$

This specification seeks to improve both the issue related to the location parameters as well as the need to ensure moments for mWTP estimates. More importantly, it ensures that the maximum expected value for $\rho$ is close to $-e^{\mu_{price}}$ rather than being close to zero as is the case for the 2LN distribution. The differences between the $\mu$LN distribution and the 2LN distribution are further described in Table 1 below.

Table 1 indicates that the mean of the 2LN distribution is different from the mean of the $\mu$LN distribution, as $-e^{\mu}$ is introduced as a shift. The probability density function (P.D.F.) $f_{\rho}(p;\mu,\sigma)$ for the 2LN distribution (where $p<0$) supports the hypothesis that decision makers can be almost completely price insensitive. Although this is also true for the $\mu$LN distribution, we argue that the main difference between the two distributions is that the point of global maximum of the P.D.F. (i.e., the mode) of the $\mu$LN distribution does not mechanically increase towards zero when $\sigma$ increases, whereas in the case of the 2LN distribution it does. This is also illustrated in Figure 1.

Figure 1 shows that when $\sigma$ increases with respect to $\mu$ (which is fixed at 0 in this example), the mode of the 2LN distribution progressively shifts toward zero, which is undesirable for deriving welfare estimates as this increases the chances of exploding ratio issues. Even when the mode is not near zero, we observe that the probability of observing values near zero is not null and, again, increases when $\sigma$ increases. This is not the case for the $\mu$LN distribution as previously stated, for which the maximum value corresponds approximately to the shift, that is, $-e^{\mu}$. In the context of this paper, this means that a mixed MNL model for which the price attribute (or any other numeraire) is assumed to follow a $\mu$LN distribution will not lead to unreasonably large WTP estimates unless the overall sensitivity to price is low. This is different from the 2LN distribution where exploding ratio issues will also arise and are in fact far more likely when the sensitivity of decision makers to the price attribute is heterogeneous.

From a theoretical standpoint, the $\mu$LN distribution has desirable properties for avoiding exploding ratio issues when deriving welfare measures from mixed MNL models. Its usefulness in practice is demonstrated by comparing its performances against other common models on five different datasets.

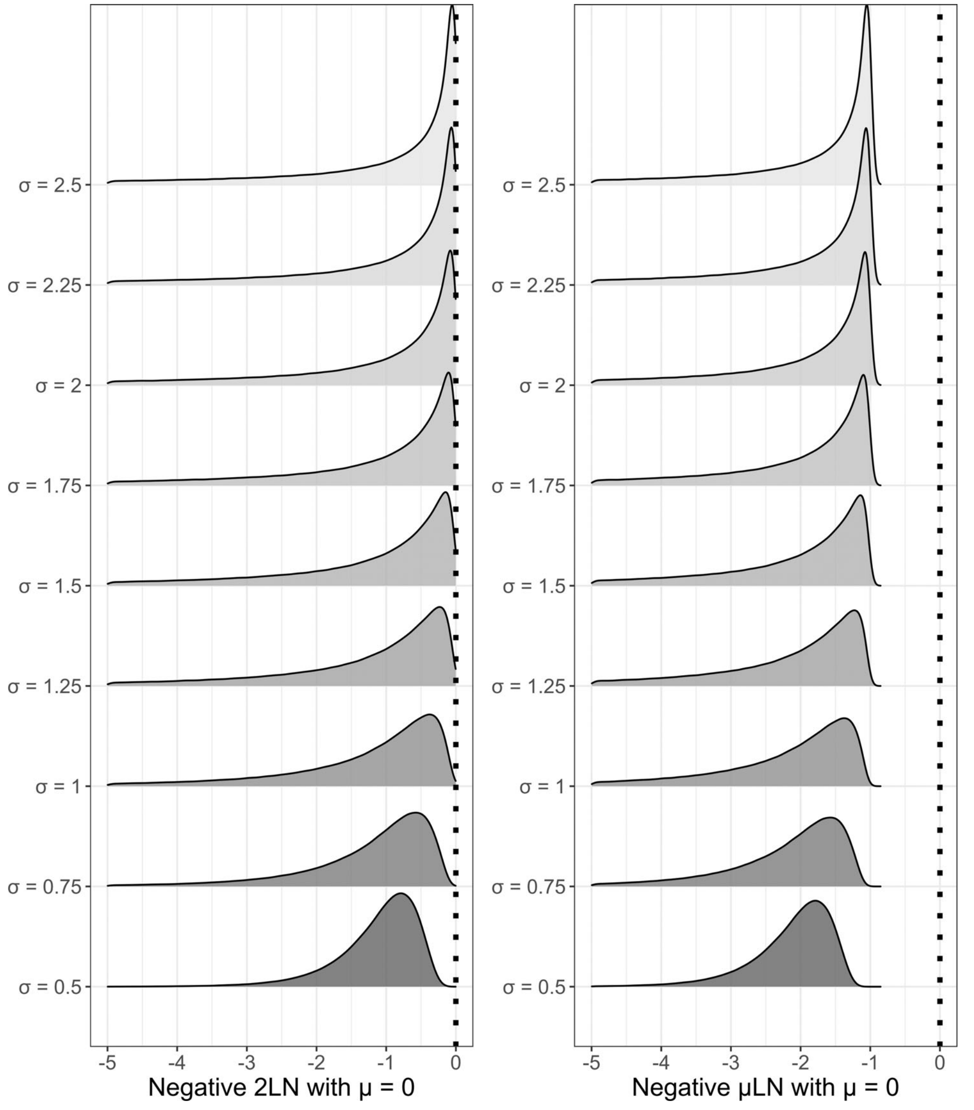


**FIGURE 1** Negative 2LN versus negative $\mu$LN distribution. Comparison of the 2LN and $\mu$LN distributions for various values of $\sigma$ (standard deviation). LN, log-normal.

# 3 | DATA

We evaluate the new shifted distribution introduced in this paper by comparing how it performs with respect to other existing model specifications using five datasets. We first introduce the five datasets, discuss the purpose served by each of them for assessing the merits of the $\mu$LN approach and finally present the criteria used to compare models and the different model specifications considered.

## 3.1 | Data Sources

The datasets featured in this paper come from four different countries (France, Poland, UK, and USA). The SP surveys vary in terms of design (number of attributes and number of choice

TABLE 2 Overview of data

| Survey | Choice tasks | Respondents | Attributes | Levels |
|---|---|---|---|---|
| Coffee | 7 | 288 | Card pay—card payment allowed | 0, 1 |
| | | | Organic—organic coffee | 0, 1 |
| | | | Fairtrade—fairtrade coffee | 0, 1 |
| | | | Recycle—cups can be recycled | 0, 1 |
| | | | Price (in €) | 0.40, 0.45, 0.50, 0.55 |
| Flood | 6 | 416 | Agri—agricultural practices against floods | 0, 1 |
| | | | Infra—protections against floods | 0, 1 |
| | | | Com—communication against floods | 0, 1 |
| | | | Price (in €) | 0, 12.5, 25, 32.5 |
| Forest | 12 | 1202 | Cen—naturalness of commercial forest | 0, 1 |
| | | | Gos—naturalness of second growth forest | 0, 1 |
| | | | Vis1—small restrictions on visitors | 0, 1 |
| | | | Vis2—large restrictions on visitors | 0, 1 |
| | | | Fee (in zl) | 25, 50, 75, 100 |
| Chicken | 12 | 816 | Test—food testing | 0, 1 |
| | | | Trace—traceability | 0, 1 |
| | | | Well—animal health | 0, 1 |
| | | | Origin—EU, Ireland or G-B | 0, 1, 2 |
| | | | Price (in GBP) | 2, 2.5, 3, 3.5, 4, 4.5 |
| Car | Up to 15 | 500 | Engine type—gas, electric (EV) or hybrid | 0, 1, 2 |
| | | | Operating cost (in $ per month) | From 2.51 to 72.29 |
| | | | Performance | 0, 1, 2 |
| | | | Range (for EV, in miles) | Up to 200 |
| | | | Body type (10 types) | 0 to 9 |
| | | | Purchase price (in $) | From 7018 to 97,301 |

*Note*: Main features of the five stated choice survey datasets used for comparing competing model specifications.

scenarios). Some SP surveys address market goods (car choice, chicken meat choice, coffee choice) whereas others address non-market goods (forest protection, erosive run-off mitigation). Two criteria were used for selecting datasets. First, each dataset has been used in past published work, in an attempt to ensure that the data used for comparing models is of a good standard according to peers. Second, some of the datasets were chosen based on the fact that they feature a reasonable number of attributes (up to 5) in order to ensure computational tractability when estimating models with correlated random parameters, which can become an issue when a full Cholesky decomposition is assumed. Each dataset is described by a keyword and a corresponding reference:

- Coffee (Sandorf et al., 2018).
- Flood (Crastes dit Sourd et al., 2014).
- Forest (Bartczak, 2015).
- Chicken (Campbell & Doherty, 2013).
- Car (Train & Weeks, 2005).

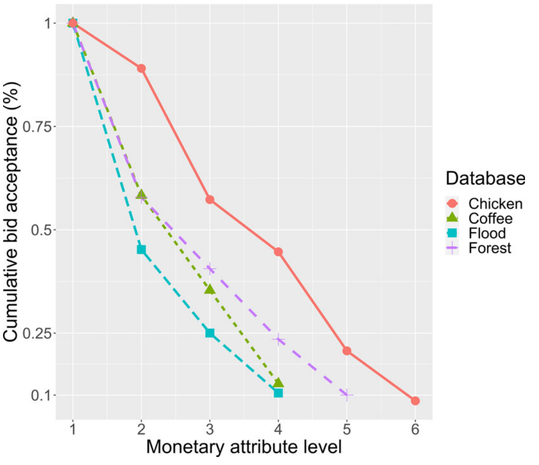


**FIGURE 2** Bid acceptance curves. Each curve reports, for a given dataset, the proportion of alternatives chosen by respondents given the level of the monetary attribute, arranged in ascending order.

The datasets are further described in Table 2 below.

The first case study is the Flood survey, which makes use of data from a SP survey conducted in France for measuring preferences for mitigating the risks related to erosive run-off events. The second case study (Coffee) is a choice experiment about preferences for different coffee machines at the University of Nantes. The third case study, labeled as Forest uses attributes describing changes in the management and quality of the Bialowieza forest in Poland. The fourth case study (Chicken) is a survey on preferences for value-added services to chicken meat in the UK. The final case study, labeled as the Car survey, investigates choices between different vehicles including electric and hybrid vehicles. The data has been collected in the USA. It is worth noting that this dataset has also been used by Train and Weeks (2005) for introducing the WTP-space approach and comparing it to the preference space approach. All five SP surveys feature three alternatives and include one status quo alternative except for the *Car* dataset, which does not. No information is available on protest answers for all the datasets except for the Flood dataset, for which 184 respondents have been removed from the original sample (619 respondents) either because they provided incomplete answers or because they were labeled as protesters. In addition, one choice task was removed from the Coffee dataset (which originally featured eight choice tasks instead of seven) because of a dominating choice, which was originally introduced for other research purposes.

## 3.2 | Sensitivity for the monetary attribute

Further information about each dataset can be derived from the analysis of the choices made by respondents as a function of the level of the monetary attribute. More precisely, and given the central importance given to the monetary attribute and its specification in the current paper, we propose to investigate whether alternatives for which the price is higher are less likely to be chosen; for most normal goods, alternatives with a higher price should be less appealing than alternatives with a lower price. Glenk et al. (2019), citing Mørkbak et al. (2010) as well as Guy and Willis (1999), report that the highest level of the monetary vector is often chosen following a rule of thumb that the alternatives with the highest level for the monetary attribute should not be selected in more than 5% to 10% of the choice situations where it is present. Although the evidence for such a rule is lacking, it is clear that the amounts presented to the respondents must be credible (Johnston et al., 2017) and that datasets for which the chosen alternatives are more often the alternatives for which the monetary attribute level is the highest suggest the presence of issues with the experimental design and/or with the engagement of respondents. We follow Glenk et al. (2019) and plot “bid acceptance curves” for each dataset, which show the percentage of alternatives chosen depending on the magnitude of the cost attribute[2] and are reported in Figure 2.

For all datasets, the highest monetary bid has been selected between 6% and 12% of the time. The bid acceptance curves appear to be "well-behaved" in all cases. This confirms that the datasets are suitable for analyzing the effect of a new distribution for the monetary parameter on various modeling outcomes introduced below. The framework for empirical testing is introduced in the next section.

## 4 | FRAMEWORK FOR EMPIRICAL TESTING

We use the Flood, Coffee, Forest, and Chicken datasets to assess the usefulness of the μLN distribution for modeling data coming from stated choice surveys. They are labeled as the core datasets in the remainder of this paper. The Car dataset features two monetary attributes, or numeraires (purchase price and operating cost) and is treated as a special case to demonstrate the usefulness of the μLN distribution also when estimating models in WTP-space. Competing models introduced below are compared based on the recommendations of Mariel, Hoyos, et al. (2021), who set guidelines for comparing models estimated on stated choice data (see Chapter 8 and 9 as well as Train and Weeks [2005] and Scarpa et al. [2008] who use a similar approach for comparing models estimated in preference and WTP space).

### 4.1 | Criteria for comparing models

Three criteria are used to evaluate and compare models: goodness of fit, moments and quantiles of mWTP distributions, and cross-validation performances.

#### Goodness of fit versus mean welfare estimates

We first compare the impact of distributional assumptions for the price attribute on model fit by comparing the log likelihood as well as the Akaike information criterion (AIC) (in Appendix) and Bayesian information criterion (BIC) for each competing model. For the models estimated in preference space, we compare whether potential increases in goodness of fit lead to exploding ratio issues, leading to unreasonable mean mWTP estimates compared to the other models considered.

#### Moments and quantiles of mWTP distributions

Second, we compare the difference between the moments and quantiles of the mWTP distributions derived from each competing model specification for all the attributes and datasets. Confidence intervals for mean mWTP estimates, calculated using standard errors of model coefficients, are computed using the Krinsky and Robb (1986, 1990) procedure.[3] We subsequently use the complete combinatorial approach proposed by Poe et al. (2005) to test whether mean mWTP measures are significantly different across models for the same attribute and dataset (see De-Magistris et al., 2016, for a similar use of the Poe test). Standard deviations and quantiles are reported in the Appendix S1.

[2]Results for the car choice dataset were not plotted because the number of levels for the monetary attribute was found to be too large. Detailed results for this dataset are available from the author upon request

[3]For each estimated model, we use the following process: 10,000 draws are taken from a multivariate normal distribution with means corresponding to the model coefficient and covariance given by the robust variance–covariance matrix of the parameter estimates of the coefficients. Based on 10,000 draws taken from the joint distribution of the coefficients, 10,000 mWTP distributions are simulated for each non-monetary attribute and model (Hole, 2007), thus giving the empirical distribution of the mean mWTP for each attribute considered allowing confidence intervals to be calculated.

### Cross-validation (out-of-sample fit)

The third and final criterion relates to cross-validation performances. This is a common method to investigate whether a given model is prone to overfitting (or, in the current case, compare whether different competing models are more or less prone to overfitting). Although it is acknowledged that the main focus in non-market valuation is on the welfare measures and not on out-of-sample fit, Mariel, Hoyos, et al. (2021) state that "*if the alternatives are assigned to a specific environmental programme or action, the individual predictions can be relevant to identify appropriate policies*." Moreover, Parady et al. (2021) recently argued that policy inference analysis should provide evidence on the generalizability (defined as the ability of a model to maintain its predictive accuracy in a different sample) of such inference. Each dataset is split into two parts: an estimation sample and a validation sample, so 50% of the respondents are allocated to the estimation sample and the remaining 50% are allocated to the validation sample. This is repeated 30 times with different, randomly selected respondents each time. The 30 samples are the same across the different model specifications for a given dataset. For each estimation sample, the resulting parameters are used to compute the log likelihood of the model on the estimation sample as well as on the corresponding validation sample. Models are evaluated by comparing the average percentage difference between the log likelihood for the validation sample and the log likelihood for the estimation sample across the 30 runs.

This protocol is similar to the protocol proposed by Train and Weeks (2005), where the sampled respondents are divided into two equal-sized subsamples and the log likelihood of the estimated models are evaluated on the other subsample. However, our protocol is very different from the approach of Sonnier et al. (2007) who included all but one choice situation for each respondent in their estimation sample and computed the log likelihood of the competing models on the single choice situation remaining for each respondent. In all cases, the objective is the same: to cross-validate the models and to compare whether different specifications are better or worse at fitting "hold-out" choice situations. Discrepancies between in and out-of-sample fit measures can be signs that a given specification is more prone to overfitting with respect to other competing models.

An important clarification at this stage is to point out that the "best" model across the specifications considered within this framework for empirical testing is not necessarily the model that yields the best goodness of fit as this can be at the expense of behavioral realism (unreasonably high WTP estimates) or generalizability as previously discussed. This is also true for the second criterion: a model that delivers plausible welfare estimates based on expert knowledge and a comparison with the other competing models is not necessarily the right model if it fits the data much worse than alternative specifications. The criteria considered here need to be weighed against each other to establish which models offer an acceptable trade-off between goodness of fit and reasonable welfare estimates.

## 4.2 | Modeling work

For each core dataset, 12 different models are estimated. The models featuring random parameters are all estimated using 2500 Sobol draws. Although this is a lower number than Czajkowski and Budziński (2019) have recently advocated, this does not substantially affect our findings[4] and is necessary to maintain a reasonable computational burden and perform a large number of model estimations and comparisons in order to provide robust empirical evidence about the usefulness of the $\mu$LN distribution. For all the *core* datasets, all the non-

[4]A series of models have been estimated again with 5000 and 10,000 Sobol draws. No substantial differences were observed. All the models have been estimated using starting values derived from the corresponding MNL model.

TABLE 3 Model specifications

| Model | Monetary attribute parameterization | Additional details |
|---|---|---|
| MNL | $\rho_{price} = \mu_{price}$ | No random heterogeneity |
| FP | $\rho_{price} = \mu_{price}$ | No random heterogeneity for price |
| LU | $\rho_{price} = -e^{\left(\mu_{price} + \sigma_{price}\zeta_{price}\right)}$ | $\zeta \sim U(0, 1)$ |
| 3LN | $\rho_{price} = \kappa - e^{\left(\mu_{price} + \sigma_{price}\zeta_{price}\right)}$ | $\zeta \sim N(0, 1)$ |
| 2LN and 2LN-c | $\rho_{price} = -e^{\left(\mu_{price} + \sigma_{price}\zeta_{price}\right)}$ | $\zeta \sim N(0, 1)$ |
| $\mu$LN and $\mu$LN-c | $\rho_{price} = -e^{\left(\mu_{price}\right)} - e^{\left(\mu_{price} + \sigma_{price}\zeta_{price}\right)}$ | $\zeta \sim N(0, 1)$ |
| WTPS and WTPS-c | $\rho_{price} = -e^{\left(\mu_{price} + \sigma_{price}\zeta_{price}\right)}$ | $\zeta \sim N(0, 1)$, see Equation (5) |
| WTPS2 and WTPS2-c | $\rho_{price} = -e^{\left(\mu_{price} + \sigma_{price}\zeta_{price}\right)}$ | $\zeta \sim N(0, 1)$, see Equations (5) and (6) |

*Note*: summary of the 12 different model specifications considered in this research.
Abbreviations: MNL, Multinomial logit; LU, log-uniform; FP, fixed price coefficient; WTPS, WTP-space.

monetary attributes have been specified using Normal draws (apart from the MNL model introduced below for which no random heterogeneity is assumed). Some of the models estimated in WTP space use flexible distributions as described in Equation (6). All models were estimated using the Apollo package for R (Hess & Palma, 2019). The different specifications considered are introduced in Table 3 below.

The MNL specification is introduced as a benchmark when it comes to low goodness of fit, as random heterogeneity in preferences is expected in all cases and also because it is still used for welfare analysis in some contexts. The fixed price coefficient specification refers to a model where all the parameters are randomly distributed apart from the monetary (Price) attribute. A negative log-uniform (LU) specification is also considered. The 3LN specification assumes that the monetary attribute follows a three-parameter negative log-normal distribution. The shift parameter labeled as $\kappa$ is estimated without constraints to verify whether mean mWTP can be derived from the model (which is not possible if $\kappa$ is found to be positive). The 2LN model refers to a model where the monetary attribute is assumed to be distributed negative log-normal while it is assumed to follow a negative $\mu$LN distribution for the model labeled as $\mu$LN. The WTP-space model is simply the counterpart of the 2LN model estimated in WTP space and the WTPS2 model is an extension of the WTPS model where each non-monetary attribute is specified as a second order polynomial of a standard normal random variable. Four models in Table 3 are estimated with and without assuming correlation between the random parameters, and the models estimated with correlated random parameters are labeled using the suffix "-c". A full Cholesky decomposition is assumed for these models (all the random parameters are correlated with one another). The ASCs, with two for each model given that all datasets feature three alternatives, are also specified as randomly distributed whenever applicable. Results for the models estimated with correlated random parameters are reported and discussed in the next section, and results for the other models are available in the Appendix S1.

Different models are estimated for the Car dataset. Given that the dataset features two monetary attributes, four models are estimated. Two models are estimated in purchase price space and where operating cost is either assumed to follow a 2LN or a $\mu$LN distribution. Two other models are estimated in operating cost space with similar distributional assumptions for purchase price. In all cases, the attribute used for normalizing the model is assumed to follow a 2LN distribution, as commonly found in the literature. The analysis of this dataset compares the goodness of fit of the four models and measures differences in terms of mean WTP estimates across models.

# 5 | RESULTS

## 5.1 | Goodness-of-fit and mean mWTP

With each of the core datasets, none of the models assuming a 3LN for the monetary parameter yielded moments for the mWTP distributions. Indeed, in all cases, the shift parameter was found to be positive. The models suffer from the same issues as would be found with a model specified with a normally distributed monetary parameter. They demonstrate better goodness-of-fit measures than the other models, but they violate basic microeconomic assumptions and cannot provide moments for mWTP estimates.

The results for the models with correlated random parameters are found in Figure 3 below, showing the BIC score for each model as well as differences in terms of mean mWTP with respect to the WTPS2-c model for each dataset. The WTPS2-c model has been chosen as a benchmark because results show, for both models with correlated and uncorrelated random parameters, evidence that assuming flexible mixture distributions is found to yield a better fit than assuming normally distributed mWTP for the majority of models. We maintain the WTPS2-c specification as a benchmark even in the cases (Forest and Chicken datasets) where the WTPS2-c model is outperformed by its corresponding WTPS-c model in terms of BIC because the AIC and log-likelihood measures are still found to be better.[5] Finally, we argue that the WTPS2-c specification is a suitable benchmark to evaluate whether the welfare measures (especially mean mWTP estimates) derived from the $\mu$LN approach are reasonable, as it is not prone to suffer from the exploding ratio issue, given that it is estimated in WTP space, while at the same time being the best fitting model in its category as discussed above.

Overall, the best performing model in terms of goodness of fit is found to be the 2LN-c model in all four cases. With regard to the Coffee dataset, although the WTPS2-c model is found to have a marginally better log likelihood, it is outperformed in terms of AIC and BIC. However, all of the 2LN-c models suffer from the exploding ratio problem, although to different degrees. The mean mWTP values derived from these models are much larger than those derived from the benchmark specification (WTPS2-c), as well as the other models, with values up to +1311% for the Infra attribute (Flood dataset). For each of the four datasets, the difference in terms of goodness of fit between the 2LN-c model and the $\mu$LN-c model is found to be marginal. The smallest difference is found for the Coffee dataset where the BIC of the 2LN-c model is 3061.36, whereas it is 3063.4 for the $\mu$LN-c model. The largest difference is found for the *Forest* dataset (20934.25 vs. 20962.07). At the same time, the $\mu$LN-c models are found to provide welfare estimates that are much closer to the benchmark model (WTPS2-c) than what is found for the 2LN-c models (+7.49% on average although there are substantial differences across datasets as reported in Figure 3). These results indicate that using a $\mu$LN distribution instead of a 2LN distribution for the cost attribute in mixed logit models allows us to mitigate the exploding ratio problem while preserving goodness of fit. Models where random parameters have not been specified as correlated feature a similar pattern of results, although the differences, both in terms of goodness of fit and welfare estimates, are larger across models.

With regard to the results for the Car dataset, as previously discussed, the SP survey features two monetary attributes (purchase price and operating cost). Hence, models estimated in WTP space can be either specified in purchase price space or in operating cost space. Detailed outputs shown in the Appendix S1 demonstrate that models estimated in WTP space are only yielding reasonable welfare estimates for the numeraire used to normalize the model, whereas the welfare estimates derived for other 2LN-distributed numeraires are substantially larger; for example, the mean mWTP measures for operating cost are 100% higher on average. Using a $\mu$LN distribution instead mitigates this issue.

[5]In addition and for each dataset, a series of likelihood ratio tests have been performed to ensure that the WTPS2 models (with and without correlated random parameters) outperform the WTPS models (with $p$-value < 0.05) in all cases.

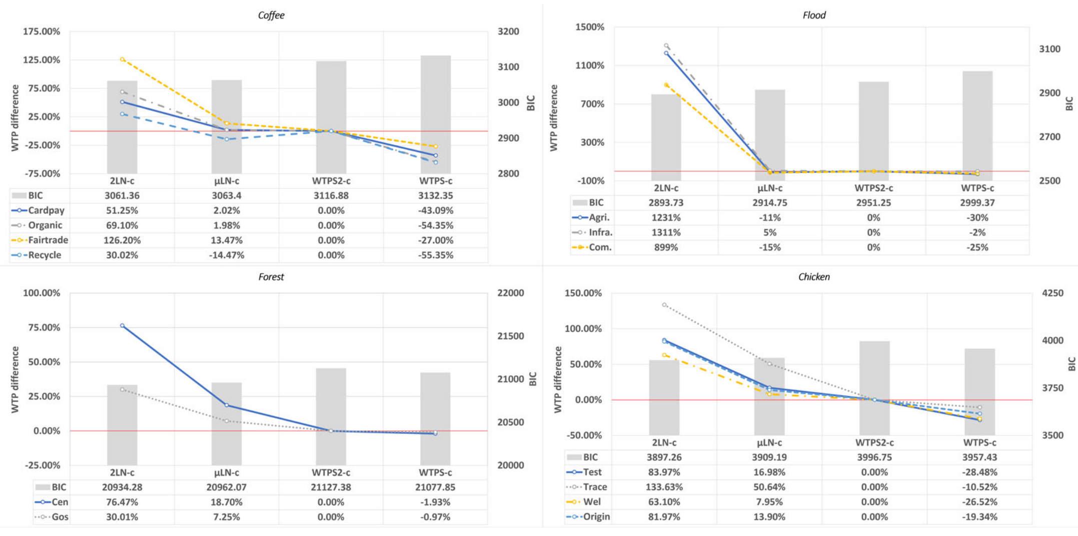


**FIGURE 3** Welfare versus fit—correlated parameters. Model performances expressed in terms of BIC and marginal willingness-to-pay (mWTP) differences (measured in percentage) with respect to the benchmark model for each *core* dataset. All models feature a full Cholesky decomposition.

All models have been found to fit the data well, although the models for which both numeraires are specified as 2LN are found to fit the data marginally better, just as has been found with the other core datasets.

Altogether, these results suggest that the $\mu$LN specification and the WTPS2 specification can converge to achieve similar results in terms of mWTP measures and goodness of fit, although the WTPS2 specification is less parsimonious in parameters. We further explore whether the $\mu$LN-c, WTPS2-c and 2LN-c specifications yield significantly different mean mWTP measures using a series of Poe tests.

## 5.2 | Moments and quantiles of welfare estimates

We test whether the mean mWTP measures derived from the 2LN-c, and WTPS2-c models are significantly different using a series of Poe tests. The WTPS-c models are not considered because they have been found to be outperformed by the WTPS2-c models as previously discussed. The confidence intervals for the mean mWTP values are further detailed in the Appendix S1 together with other moments and quantiles. Results for mean mWTP measures are reported in Table 4 below.

The null hypothesis of the test is that mWTP estimates are equal between two competing model specifications for a given attribute. Hence, results for Table 4 should be interpreted as follows: a $p$-value smaller than 0.05 or larger than 0.95 indicates that the mean mWTP values are different at the 5% level of confidence, whereas NA indicates that the mean mWTP estimates were not found to be significantly different from zero at the 5% level, in which case no test was performed. This is similar to how Sandorf et al. (2020) have interpreted the Poe test.[6]

Results indicate that in all cases, the hypothesis that the mean mWTP values derived from the $\mu$LN-c models are equal to the values derived from the WTPS2-c models cannot be rejected at the 5% level. For the Chicken dataset, the mean mWTP for the Trace attribute is found to be significantly different at the 10% level. The values derived from the 2LN-c model for the Flood dataset are

[6]The Poe test itself was conducted using the cmdlR package in R from the same author (Sandorf, 2020).

TABLE 4 Poe test $p$-values on differences between mean marginal willingness-to-pay (WTP) across models

| Dataset | Attribute | $\mu$LN-c = WTPS2-c | 2LN-c = $\mu$LN-c | LN-c = WTPS2-c |
|---|---|---|---|---|
| Coffee | Card pay | 0.4521 | 0.0823 | 0.0489 |
| | Organic | 0.4771 | 0.0862 | 0.0534 |
| | Fairtrade | 0.3392 | 0.0700 | 0.0339 |
| | Recycle | 0.7657 | 0.1355 | 0.2253 |
| Flood | Agri | 0.7340 | 0.0000 | 0.0000 |
| | Infra | 0.4001 | 0.0000 | 0.0000 |
| | Com | 0.8502 | 0.0003 | 0.0003 |
| Forest | Cen | 0.1778 | 0.1762 | 0.0900 |
| | Gos | 0.4001 | 0.4082 | 0.3713 |
| | Vis1 | NA | NA | NA |
| | Vis2 | NA | NA | NA |
| Chicken | Test | 0.2817 | 0.0681 | 0.0277 |
| | Trace | 0.0887 | 0.1367 | 0.0210 |
| | Wel | 0.3718 | 0.1079 | 0.0659 |
| | Origin | 0.2892 | 0.0641 | 0.0193 |

*Note*: $p$-values larger than 0.95 or smaller than 0.05 indicate significantly different mean mWTP at the 5% level of confidence.

found to be significantly different from the mWTP values derived from the $\mu$LN-c and WTPS2-c models at the 1% level of confidence. Results are more nuanced for the three other datasets. For the Coffee and Chicken datasets, most of the mean mWTP values derived from the $\mu$LN-c models are only found to be significantly different from the values derived from the 2LN-c models at the 10% level, although these differences are significant at the 5% level for the WTPS2-c models. Finally, results for the Forest dataset show that mean mWTP estimates are not significantly different across the three models except for the Cen attribute where the mean mWTP derived from the WTPS2-c model is found to be significantly lower than the value derived from the 2LN-c model at the 10% level of confidence.

Overall, these results suggest that using a $\mu$LN distribution for the monetary attribute for models estimated in preference space yields mean mWTP estimates, which are not significantly different than the values derived from models estimated in WTP space with flexible mWTP distributions. At the same time, the $\mu$LN distribution avoids the potential exploding ratio problem, which can be occasionally encountered when using a 2LN distribution for the monetary attribute.

## 5.3 | Cross-validation results

The final criterion for evaluating the usefulness of the $\mu$LN distribution relates to cross-validation performances. Using the four *core* datasets, we compare the 2LN-c, $\mu$LN-c, WTPS-c and WTPS2-c specifications. Results for the models with correlated random parameters are reported in Table 5 below.

Results should be interpreted as follows: For the Coffee dataset, the specification which yields the best mean percentage difference (MPD) between out-of-sample and in-sample fit is found to be the $\mu$LN-c model. Based on 30 runs, the log likelihood for the validation sample is found to be 5.69% lower than the log likelihood for the estimation sample on average. The $\mu$LN-c specification yields the higher MPD in three cases out of four (Coffee, Flood, and Chicken dataset). The MPD measures for the 2LN-c and $\mu$LN-c specifications are found to be nearly the same for the Forest dataset.

TABLE 5 Cross validation: % difference across 30 runs

| Dataset | Model | Mean | Standard deviation | Minimum | Maximum |
|---|---|---|---|---|---|
| Coffee | 2LN-c | −6.04 | 5.35 | −16.50 | +7.73 |
| | $\mu$LN-c | −5.69 | 5.21 | −15.18 | +7.96 |
| | WTPS-c | −6.20 | 5.18 | −17.13 | +6.56 |
| | WTPS2-c | −7.96 | 5.65 | −16.46 | +5.93 |
| Flood | 2LN-c | −5.11 | 6.86 | −22.56 | +8.34 |
| | $\mu$LN-c | −5.03 | 7.04 | −23.84 | +8.60 |
| | WTPS-c | −5.87 | 6.96 | −22.77 | +8.33 |
| | WTPS2-c | −5.60 | 6.81 | −22.49 | +7.18 |
| Forest | 2LN-c | −0.99 | 1.95 | −4.70 | +3.46 |
| | $\mu$LN-c | −0.99 | 1.97 | −4.79 | +3.47 |
| | WTPS-c | −1.24 | 1.90 | −4.53 | +2.94 |
| | WTPS2-c | −1.60 | 1.94 | −5.22 | +2.62 |
| Chicken | 2LN-c | −3.07 | 4.17 | −10.71 | +5.16 |
| | $\mu$LN-c | −2.88 | 4.10 | −10.73 | +5.21 |
| | WTPS-c | −3.26 | 3.88 | −10.11 | +6.10 |
| | WTPS2-c | −4.13 | 4.00 | −11.71 | +4.73 |

*Note*: percentage difference between out-of-sample and in-sample fit for each specification and dataset.

These two specifications also outperform the models estimated in WTP space in all cases. The lowest MPD are found for the WTPS2-c specification except for the Flood dataset where the worst specification is the WTPS-c model.

These findings suggest that using nonparametric mixtures of distributions for modeling random heterogeneity in preferences is, unsurprisingly, more prone to overfitting issues than parametric distributions, although the extent of the issue is very moderate for all models and datasets. Again, the $\mu$LN distribution is not found to feature any significant disadvantage compared to the 2LN distribution while also providing much more reasonable welfare estimates as previously stated. A similar pattern of results is found for the models estimated without correlated random parameters, although MPD measures are found to be generally higher.

This section has demonstrated that assuming a $\mu$LN distribution for the monetary attribute leads to welfare estimates that are more reasonable than when assuming a 2LN distribution while at the same time being more parsimonious in parameters than models estimated in WTP space with flexible mixture distributions. The $\mu$LN specification has also been found to have higher cross-validation performances than other competing models.

## 6 | DISCUSSION AND CONCLUSIONS

Mixed MNL models are the most widely used specification for fitting stated choice experiments data in non-market valuation, transport, and health, among other fields (Mahieu et al., 2017). The past 20 years have seen a profusion of competing specifications, some of which aim at fitting the data better, whereas others seek to provide more reasonable distributions of mWTP. However, in some cases, improving the fit of a given model leads to implausibly large welfare estimates, and mitigating this issue leads, in turn, to a decrease in the fit of the model. This trade-off, largely illustrated by the opposition between preference space and WTP-space models, has led practitioners to suggest finding new modeling strategies for fitting the data better in WTP space and/or mitigating the so-called exploding ratio issue in preference space (Train & Weeks, 2005).

In this paper, we have investigated the usefulness of shifted negative log-normal distributions for the price attribute in mixed logit models. In addition to evaluating the traditional shifted negative log-normal distribution, labeled as the three-parameter log-normal and recently studied by McFadden (2022), we have introduced a new parameterization labeled as the $\mu$LN distribution. These distributions have been compared to more common modeling strategies found in the literature. These comparisons have been made following a framework for empirical testing which considered goodness of fit, mean marginal WTP measures and out-of-sample fit.

Although the 3LN has been found to be problematic due to the fact that it cannot necessarily ensure the existence of moments for the mWTP distributions, using a $\mu$LN distribution for the monetary attribute has been found to deliver welfare estimates, which are not significantly different from the values derived from a model estimated in WTP space with correlated parameters and flexible distributions for modeling taste heterogeneity. Moreover, the $\mu$LN approach delivers welfare estimates that are substantially lower than those derived from a model where the price attribute is assumed to follow a 2LN distribution. At the same time, the $\mu$LN approach fits the data almost as well as the 2LN distribution (the models are virtually equivalent in most cases). In comparison, the WTP-space approach is outperformed in terms of goodness of fit, sometimes considerably, in most of the cases considered. As previously stated, improving the flexibility of the WTP-space model by introducing nonparametric mixtures of normal distributions reduces the gap between the $\mu$LN and the WTP-space specifications in terms of goodness of fit and welfare measures. However, that is not all an analyst should be concerned with. Convenience and behavioral realism can also be reasons to consider the $\mu$LN distribution.

Models specified in WTP space generally assume that the monetary parameter is log-normally distributed. This has strong implications in terms of behavior as this implicitly assumes that some respondents are making their choices solely based on random factors, from the point of view of the analyst, given that the 2LN distribution can reach values that are very close to zero. Although this can be true in some settings (see e.g., Sandorf et al., 2020), this is a modeling assumption that can be harder to defend when using revealed preference data as this would imply that some people make real-world choices completely randomly. This makes the $\mu$LN distribution a good specification for the monetary attribute in WTP space even when there is only one numeraire, as this prevents the monetary (scale) parameter being close to zero and hence does not assume that observed choices can be completely random. This assumption is not only reasonable when using revealed preference data but also stated preference data as it can prevent the random choices respondents make in surveys affecting the real-world implications of the analysis.

It is also worth noting that although models specified in WTP space have the advantage of allowing the analyst to directly specify and control WTP distributions rather than deriving them from marginal utility distributions, they are also more prone to suffer from estimation problems. This is because "*parameters enter nonlinearly in the model in WTP-space*" (Scarpa et al., 2008), which can lead to convergence issues (Mariel, Hoyos, et al., 2021). How pervasive convergence issues are in WTP space has not been systematically documented in the literature, although some authors have reported issues, for example, Toledo-Gallegos et al. (2021). Such complications are more likely to arise when estimating models featuring nonparametric mixtures of distributions (such as the models labeled as *WTPS2* in our paper) because more flexible models can be more difficult to estimate, with convergence issues and problems with parameter significance (Hess, 2010). The fact that these issues exist does not rule out the use of models specified in WTP space as they can sometimes outperform their counterpart in preference space, as previously discussed, in addition to offering more control over mWTP distributions. Similarly, the fact that models estimated in preference space with a 2LN distributed monetary parameter can potentially lead to implausibly large welfare estimates does not mean that this is bound to happen regardless of the data at hand.

Nevertheless, we make the case that the $\mu$LN distribution is more convenient for exploratory analysis of choice data as it is less prone to suffer from estimation issues than a model estimated in WTP space, while at the same time fitting the data well and yielding reasonable welfare estimates.

Despite the large amount of evidence provided in this paper, it is possible that our results are context specific. More precisely, most of the non-monetary attributes featured in the five datasets considered have been specified as normally distributed. Different results could have been found if more datasets from the transport literature had been considered because typical attributes such as travel time are often specified as negative log normal. The ratio of a 2LN distribution and a $\mu$LN distribution might yield different results than the ratio of a normally distributed coefficient and a $\mu$LN-distributed coefficient. Future research on this topic will hence investigate the usefulness of the proposed method in different empirical contexts and propose a Monte-Carlo evaluation of the $\mu$LN distribution, similarly to what Scarpa et al. (2021) have recently proposed for logit-mixed logit models.

## ACKNOWLEDGMENTS

I am grateful for the valuable comments made by Olivier Beaumais, Danny Campbell, Mikołaj Czajkowski, Thijs Dekker, Stephane Hess, Jürgen Meyerhoff, Erlend Dancke Sandorf, Kenneth Train, and for sharing data. Special thanks to the CRIANN (France) for letting me use the Myria HPC system. The opinions and any omissions in this paper remain my responsibility.

## REFERENCES

Alfnes, Frode, Atle G. Guttormsen, Gro Steine, and Kari Kolstad. 2006. "Consumers' Willingness to Pay for the Color of Salmon: A Choice Experiment with Real Economic Incentives." *American Journal of Agricultural Economics* 88(4): 1050–61.

Bartczak, Anna. 2015. "The Role of Social and Environmental Attitudes in Non-Market Valuation: An Application to the Bia lowieża Forest." *Forest Policy and Economics* 50: 357–65.

Cameron, Trudy Ann. 1988. "A New Paradigm for Valuing Non-Market Goods Using Referendum Data: Maximum Likelihood Estimation by Censored Logistic Regression." *Journal of Environmental Economics and Management* 15(3): 355–79.

Campbell, Danny, and Edel Doherty. 2013. "Combining Discrete and Continuous Mixing Distributions to Identify Niche Markets for Food." *European Review of Agricultural Economics* 40(2): 287–312.

Crastes dit Sourd, Romain, Olivier Beaumais, Ouerdia Arkoun, Dimitri Laroutis, Pierre- Alexandre Mahieu, Bénédicte Rulleau, Salima Hassani-Taibi, Vladimir Stefan Barbu, and David Gaillard. 2014. "Erosive Runoff Events in the European Union: Using Discrete Choice Experiment to Assess the Benefits of Integrated Management Policies when Preferences Are Heterogeneous." *Ecological Economics* 102: 105–12.

Czajkowski, Mikołaj, and Wiktor Budziński. 2019. "Simulation Error in Maximum Likelihood Estimation of Discrete Choice Models." *Journal of Choice Modelling* 31: 73–85.

Czine, Péter, Áron Török, Károly Pető, Péter Horváth, and Péter Balogh. 2020. "The Impact of the Food Labeling and Other Factors on Consumer Preferences Using Discrete Choice Modeling—The Example of Traditional Pork Sausage." *Nutrients* 12(6): 1768.

Daly, Andrew, Stephane Hess, and Kenneth Train. 2012. "Assuring Finite Moments for Willingness to Pay in Random Coefficient Models." *Transportation* 39(1): 19–31.

Du, Xiaodong, Hongli Feng, and David A. Hennessy. 2017. "Rationality of Choices in Subsidized Crop Insurance Markets." *American Journal of Agricultural Economics* 99(3): 732–56.

Fosgerau, Mogens, and Stefan L. Mabit. 2013. "Easy and Flexible Mixture Distributions." *Economics Letters* 120(2): 206–10.

Giergiczny, Marek, Sviataslau Valasiuk, Mikolaj Czajkowski, Maria De Salvo, and Giovanni Signorello. 2012. "Including Cost Income Ratio into Utility Function as a Way of Dealing with 'Exploding' Implicit Prices in Mixed Logit Models." *Journal of Forest Economics* 18(4): 370–80.

Glenk, Klaus, Jürgen Meyerhoff, Faical Akaichi, and Julia Martin-Ortega. 2019. "Revisiting Cost Vector Effects in Discrete Choice Experiments." *Resource and Energy Economics* 57: 135–55.

Guy, Garrod, and Kenneth G. Willis. 1999. *Economic Valuation of the Environment: Methods and Case Studies*. Cheltenham, UK: Edward Elgar.

Hess, Stephane. 2010. "Conditional Parameter Estimates from Mixed Logit Models: Distributional Assumptions and a Free Software Tool." *Journal of Choice Modelling* 3(2): 134–52.

Hess, Stephane, and David Palma. 2019. "Apollo: A Flexible, Powerful and Customisable Freeware Package for Choice Model Estimation and Application." *Journal of Choice Modelling* 32: 100170.

Hole, Arne Risa. 2007. “A Comparison of Approaches to Estimating Confidence Intervals for Willingness to Pay Measures.” *Health Economics* 16(8): 827–40.

Johnston, Robert J., Kevin J. Boyle, Wiktor Adamowicz, Jeff Bennett, Roy Brouwer, Trudy Ann Cameron, W. Michael Hanemann, et al. 2017. “Contemporary Guidance for Stated Preference Studies.” *Journal of the Association of Environmental and Resource Economists* 4(2): 319–405.

Kalecki, Michał. 1945. “On the Gibrat Distribution.” *Econometrica: Journal of the Econometric Society* 13: 161–70.

Krinsky, Itzhak, and A. Leslie Robb. 1986. “On Approximating the Statistical Properties of Elasticities.” *Review of Economics and Statistics* 68: 715–9.

Krinsky, Itzhak, and A. Leslie Robb. 1990. “On Approximating the Statistical Properties of Elasticities: A Correction.” *Review of Economics and Statistics* 72: 189–90.

De-Magistris, Tiziana, Belinda López-Galán, and Vincenzina Caputo. 2016. “The Impact of Body Image on the WTP Values for Reduced-Fat and Low-Salt Content Potato Chips among Obese and Non-Obese Consumers.” *Nutrients* 8(12): 830.

Mahieu, Pierre-Alexandre, Henrik Andersson, Olivier Beaumais, Romain Crastes dit Sourd, Stephane Hess, and François-Charles Wolff. 2017. “Stated Preferences: A Unique Database Composed of 1657 Recent Published Articles in Journals Related to Agriculture, Environment, or Health.” *Review of Agricultural, Food and Environmental Studies* 98(3): 201–20.

Mariel, Petr, Simona Demel, and Alberto Longo. 2021. “Modelling Welfare Estimates in Discrete Choice Experiments for Seaweed-Based Renewable Energy.” *PLoS One* 16(11): e0260352.

Mariel, Petr, David Hoyos, Jürgen Meyerhoff, Mikolaj Czajkowski, Thijs Dekker, Klaus Glenk, Jette Bredahl Jacobsen, et al. 2021. *Environmental Valuation with Discrete Choice Experiments: Guidance on Design, Implementation and Data Analysis*. Cham, Switzerland: Springer.

McFadden, Daniel. 2022. “Instability in Mixed Logit Demand Models.” *Journal of Choice Modelling* 43: 100353.

Meginnis, Keila, Nick Hanley, Lazaaro Mujumbusi, and Poppy H. L. Lamberton. 2020. “Nonmonetary Numeraires: Varying the Payment Vehicle in a Choice Experiment for Health Interventions in Uganda.” *Ecological Economics* 170: 106569.

Mørkbak, Morten Raun, Tove Christensen, and Dorte Gyrd-Hansen. 2010. “Choke Price Bias in Choice Experiments.” *Environmental and Resource Economics* 45(4): 537–51.

Parady, Giancarlos, David Ory, and Joan Walker. 2021. “The Overreliance on Statistical Goodness-of-Fit and Under-Reliance on Model Validation in Discrete Choice Models: A Review of Validation Practices in the Transportation Academic Literature.” *Journal of Choice Modelling* 38: 100257.

Peterson, Jeffrey M., Craig M. Smith, John C. Leatherman, Nathan P. Hendricks, and John A. Fox. 2015. “Transaction Costs in Payment for Environmental Service Contracts.” *American Journal of Agricultural Economics* 97(1): 219–38.

Poe, Gregory L., Kelly L. Giraud, and John B. Loomis. 2005. “Computational Methods for Measuring the Difference of Empirical Distributions.” *American Journal of Agricultural Economics* 87(2): 353–65.

Rigby, Dan, Kelvin Balcombe, and Michael Burton. 2009. “Mixed Logit Model Performance and Distributional Assumptions: Preferences and GM Foods.” *Environmental and Resource Economics* 42(3): 279–95.

Rungie, Cam, Riccardo Scarpa, and Mara Thiene. 2014. “The Influence of Individuals in Forming Collective Household Preferences for Water Quality.” *Journal of Environmental Economics and Management* 68(1): 161–74.

Rungie, Campbell M., Leonard V. Coote, and Jordan J. Louviere. 2011. “Structural Choice Modelling: Theory and Applications to Combining Choice Experiments.” *Journal of Choice Modelling* 4(3): 1–29.

Sandorf, Erlend Dancke. 2020. cmdlR: Choice Modeling in R, v0.0.4 https://github.com/edsandorf/cmdlR

Sandorf, Erlend Dancke, Romain Crastes dit Sourd, and Pierre-Alexandre Mahieu. 2018. “The Effect of Attribute-Alternative Matrix Displays on Preferences and Processing Strategies.” *Journal of Choice Modelling* 29: 113–32.

Sandorf, Erlend Dancke, Lars Persson, and Thomas Broberg. 2020. “Using an Integrated Choice and Latent Variable Model to Understand the Impact of “Professional” Respondents in a Stated Preference Survey.” *Resource and Energy Economics* 61: 101178.

Sangal, B. P., and Asit K. Biswas. 1970. “The 3-Parameter Lognormal Distribution and its Applications in Hydrology.” *Water Resources Research* 6(2): 505–15.

Scarpa, Riccardo, Cristiano Franceschinis, and Mara Thiene. 2021. “Logit Mixed Logit under Asymmetry and Multimodality of WTP: A Monte Carlo Evaluation.” *American Journal of Agricultural Economics* 103(2): 643–62.

Scarpa, Riccardo, Mara Thiene, and Francesco Marangon. 2008. “Using Flexible Taste Distributions to Value Collective Reputation for Environmentally Friendly Production Methods.” *Canadian Journal of Agricultural Economics/Revue Canadienne d'Agroeconomie* 56(2): 145–62.

Scarpa, Riccardo, Mara Thiene, and Kenneth Train. 2008. “Utility in Willingness to Pay Space: A Tool to Address Confounding Random Scale Effects in Destination Choice to the Alps.” *American Journal of Agricultural Economics* 90(4): 994–1010.

Sonnier, Garrett, Andrew Ainslie, and Thomas Otter. 2007. “Heterogeneity Distributions of Willingness-to-Pay in Choice Models.” *Quantitative Marketing and Economics* 5(3): 313–31.

Svenningsen, Lea S., and Jette Bredahl Jacobsen. 2018. “Testing the Effect of Changes in Elicitation Format, Payment Vehicle and Bid Range on the Hypothetical Bias for Moral Goods.” *Journal of Choice Modelling* 29: 17–32.

Thiene, Mara, and Riccardo Scarpa. 2009. “Deriving and Testing Efficient Estimates of WTP Distributions in Destination Choice Models.” *Environmental and Resource Economics* 44(3): 379–95.

Toledo-Gallegos, Valeria M., Jed Long, Danny Campbell, Tobias Börger, and Nick Hanley. 2021. "Spatial Clustering of Willingness to Pay for Ecosystem Services." *Journal of Agricultural Economics.* 72: 673–97.

Train, Kenneth. 2016. "Mixed Logit with a Flexible Mixing Distribution." *Journal of Choice Modelling* 19: 40–53.

Train, Kenneth, and Melvyn Weeks. 2005. "Discrete Choice Models in Preference Space and Willingness-to-Pay Space." In *Applications of Simulation Methods in Environmental and Resource Economics* edited by Riccardo Scarpa and Anna Alberini, 1–16. Dordrecht, the Netherlands: Springer.

Uz, Dilek, Steven Buck, and David Sunding. 2022. "Fixed or mixed? Farmer-level heterogeneity in response to changes in salinity." *American Journal of Agricultural Economics.* 104(4): 1343–63. https://doi.org/10.1111/ajae.12270.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Crastes dit Sourd, Romain. 2023. "A New Empirical Approach for Mitigating Exploding Implicit Prices in Mixed Multinomial Logit Models." *American Journal of Agricultural Economics* 1–20. https://doi.org/10.1111/ajae.12367