Journal of Choice Modelling 60 (2026) 100624

Contents lists available at ScienceDirect

Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

Full Length Article

Inference in downstream analysis using individual-level posterior means from mixed logit models☆

Danny Campbell [a,*], Erlend Dancke Sandorf [b], Tobias Börger [c], Thijs Dekker [d]

[a] Stirling Business School, University of Stirling, United Kingdom
[b] School of Economics and Business, Norwegian University of Life Sciences, Norway
[c] Department of Business and Economics, Berlin School of Economics and Law (HWR Berlin), Germany
[d] Institute for Transport Studies, University of Leeds, United Kingdom

ARTICLE INFO

Keywords:
Mixed logit
Second-stage inference
Conditional distributions
Preference heterogeneity
Bootstrap simulation

ABSTRACT

Mixed logit models are widely used to recover individual-level preferences for secondary analysis, yet both first-stage sampling uncertainty and variability in conditional distributions are often overlooked. This technical note illustrates the implications of ignoring these sources of uncertainty and provides reproducible R code, compatible with the `apollo` package, to better approximate the empirical sampling distribution and improve the reliability of second-stage inference.

1. Introduction

Mixed logit models, including random parameter models and latent class models, are used to estimate unobserved preference heterogeneity (McFadden and Train, 2000). Using post-estimation techniques, researchers routinely derive the mean of each individual's conditional (posterior) distribution as a point estimate of marginal utility or willingness-to-pay (see e.g., Revelt and Train, 2000; Train, 2003; Sillano and de Dios Ortúzar, 2005; Greene et al., 2005). Across diverse applied fields (including environmental, transport, health economics, and marketing) these conditional means are commonly deployed in downstream second-stage analyses to examine the determinants of preference heterogeneity or to segment consumer populations. In practice, they are regularly used as dependent or independent variables in regression models (e.g., Campbell, 2007; Börger et al., 2021), spatial modelling (e.g., Campbell et al., 2008, 2009; Johnston and Ramachandran, 2014; Czajkowski et al., 2016; Vij and Hess, 2025), and hypothesis testing or group comparisons (e.g., Sandorf et al., 2017; Richter and Pollitt, 2018). However, this widespread practice frequently ignores that these conditional means are: i) dependent on the estimated model parameters, inheriting the sampling variability from the first-stage estimation, and (ii) themselves point summaries of a posterior distribution and therefore subject to their own inherent uncertainty (see, e.g., Hess, 2010; Sarrias, 2020; Chappell et al., 2022). Omitting these sources of variance risks underestimating standard errors in the second-stage analysis, thereby inflating the likelihood of Type I errors. This problem directly parallels the well-known variance-propagation issues in classic two-stage econometric modelling (e.g., Heckman, 1979; Newey, 1984; Pagan, 1984; Murphy and Topel, 1985).

To ensure valid inference, both the sampling variation of the first-stage model parameters and the probabilistic nature of the conditional distributions must be reflected in downstream analyses. In this note, we consider the large sample properties of estimators based on conditional means, outline how to incorporate these uncertainties using simulation-based methods and

☆ This article is part of a Special issue entitled: 'Standalone contributions' published in Journal of Choice Modelling.

* Corresponding author.
E-mail addresses: danny.campbell@stir.ac.uk (D. Campbell), erlend.dancke.sandorf@nmbu.no (E.D. Sandorf), tobias.boerger@hwr-berlin.de (T. Börger), t.dekker@leeds.ac.uk (T. Dekker).

https://doi.org/10.1016/j.jocm.2026.100624
Received 22 December 2025; Received in revised form 16 June 2026; Accepted 27 July 2026
Available online 11 August 2026

provide reproducible code compatible with the `apollo` package in R (Hess and Palma, 2025). While our illustration draws on an environmental application, the problem and solutions apply broadly to any context using conditional estimates from mixed logit models.

2. Second-stage inference with estimated posterior means

To generate the sampling distribution for the second-stage analysis, we build on previous work examining the sampling properties of individual-specific conditional estimates (e.g., see Revelt and Train, 2000; Sarrias and Daziano, 2018; Chappell et al., 2022). We depart from their work by focusing on the properties of the downstream estimators where these conditional estimates are used as inputs. We outline the key steps below emphasising the appropriate method for generating the conditional estimate's sampling distribution in a second-stage context to assess both consistency and efficiency.

2.1. Recovering individual-level posterior means

Let $\boldsymbol{\beta}_n^{\star}$ denote the true (unknown) vector of marginal utilities for individual n. In estimation, we fit a mixed logit model to a random sample of N individuals, each providing T_n choice observations to recover population-level parameters $\boldsymbol{\theta}$ that characterise the distribution of preferences in the population. Depending on the mixed logit model used, we obtain either parametric random coefficients (e.g., means and variances) or class-level utilities together with membership probabilities. To obtain the individual-level likelihood contributions ω_{nr} (conditional weights) for each individual n, we first draw R samples, denoted $\boldsymbol{\beta}_{nr}$, from the multivariate distribution defining the random coefficients in the mixed logit model. Importantly, these draws are conditional on the estimated hyperparameters of the distribution, so that uncertainty arises both from the sampling of individual-level coefficients and from the point estimates of the distributional parameters themselves.[1] For random parameters logit models, the posterior or conditional density generally does not have a closed-form solution, motivating the need for simulation. The conditional weights ω_{nr} are then constructed from these draws. In a standard random parameters logit model, the unconditional (prior) weights π_{nr} are uniform, i.e., $\pi_{nr} = R^{-1}$ for all r, whereas in a latent class model they correspond to the unconditional class membership probabilities, where R now denotes the number of classes and $\boldsymbol{\beta}_{nr}$ class-specific parameters. In a frequentist framework, ω_{nr} provides an approximate conditional posterior given the estimated hyperparameters, the observed sequence of choices $\mathbf{y}_n$, and the associated explanatory variables $\mathbf{X}_n$, and the simulated draws $\boldsymbol{\beta}_{nr}$. Each weight is computed as:

$$\omega_{nr} = \frac{\Pr\left(\mathbf{y}_n \mid \mathbf{X}_n, \boldsymbol{\beta}_{nr}\right) \pi_{nr}}{\sum_{r=1}^{R} \Pr\left(\mathbf{y}_n \mid \mathbf{x}_n, \boldsymbol{\beta}_{nr}\right) \pi_{nr}}. \tag{1}$$

The expected conditional mean for attribute k is:

$$\hat{\beta}_{nk} = \sum_{r=1}^{R} \omega_{nr} \beta_{nkr}. \tag{2}$$

This provides the posterior mean approximation for individual n, serving as an approximation of $\beta_{nk}^{\star}$. In the case of models with a continuous mixing distribution (such as the random parameters logit) the posterior lacks a closed-form expression and must be approximated via simulation, rendering the resulting conditional distribution discrete even for large R (notwithstanding that the underlying posterior is effectively continuous). By contrast, in a latent class model the posterior weights are available in closed form, as the finite mixture structure means the conditional class membership probabilities can be computed exactly given estimated parameters. Thus, for any mixed logit model, the simulated approximation to the posterior distribution of β_{nk} for individual n, conditional on the estimated population-level parameters $\hat{\boldsymbol{\theta}}$, is given by:

$$\hat{F}_{nk}\left(\beta \mid \hat{\boldsymbol{\theta}}\right) \equiv \left\{\left(\beta_{nk1}, \omega_{n1}\right), \left(\beta_{nk2}, \omega_{n2}\right), \ldots, \left(\beta_{nkR}, \omega_{nR}\right)\right\}. \tag{3}$$

From this discrete distribution, the variance and other moments can be calculated as weighted sums of powers of the outcomes, using the associated weights ω_{nr} as the probability mass function.

2.2. Asymptotic properties of the second-stage estimator

Once the conditional mean $\hat{\beta}_{nk}$ is obtained, it can be used in a second-stage analysis. We focus on ordinary least squares (OLS) regression analysis treating $\hat{\beta}_{nk}$ as the dependent variable, though similar properties and implications generally apply across various tests and settings. This creates a two-stage estimation problem. In the first stage, a mixed logit model is estimated to obtain the population-level parameters $\hat{\boldsymbol{\theta}}$. In the second stage, the resulting individual-level preference estimates are regressed on observed characteristics. The key question is how to correctly account for the fact that $\hat{\beta}_{nk}$ is itself estimated, rather than known exactly. To

[1] In a fully Bayesian framework, sequential simulation via a Gibbs sampler would ensure convergence to the true marginal posterior (e.g., see Dekker et al., 2025), thereby capturing both sources of uncertainty. The frequentist approximation employed here conditions on estimated hyperparameters and thus treats estimation error as a separate component.

answer this it is necessary to establish the asymptotic properties of the second-stage estimators. Define the posterior mean vector for individual n as:

$$\boldsymbol{\beta}_n(\boldsymbol{\theta}) \equiv \mathbb{E}_{\boldsymbol{\theta}} \left(\boldsymbol{\beta}_n \mid \mathbf{y}_n, \mathbf{X}_n \right), \tag{4}$$

which makes explicit that $\boldsymbol{\beta}_n(\boldsymbol{\theta})$ is a function of the first-stage parameters $\boldsymbol{\theta}$ and the observed data. Note that $\boldsymbol{\beta}_n(\boldsymbol{\theta})$ denotes the vector of posterior means evaluated at the true hyperparameters. This need not equal the true individual-specific taste parameter $\boldsymbol{\beta}_n^{\star}$: the two coincide only in the limit as $T_n \to \infty$, since the consistency of individual-level posterior means requires a growing number of choice observations per individual (see Revelt and Train, 2000; Sarrias and Daziano, 2018; Chappell et al., 2022). In any finite sample, the posterior mean is shrunk toward the population mean, and the degree of shrinkage increases as T_n decreases (Sarrias, 2020). In practice, since $\boldsymbol{\theta}$ is unknown, we substitute the first-stage estimate $\hat{\boldsymbol{\theta}}$, giving the plug-in posterior mean for attribute k as:

$$\hat{\beta}_{nk} = \beta_{nk}(\hat{\boldsymbol{\theta}}), \tag{5}$$

which is what is actually used in the second stage. Because $\hat{\boldsymbol{\theta}} \neq \boldsymbol{\theta}$ in any finite sample, $\hat{\beta}_{nk} \neq \beta_{nk}(\boldsymbol{\theta})$; the plug-in posterior mean inherits the estimation error from the first stage.

Suppose the true second-stage OLS relationship is[2]:

$$\beta_{nk}(\boldsymbol{\theta}) = \mathbf{z}_n^{\top} \boldsymbol{\delta} + \varepsilon_n, \qquad \mathbb{E}\left(\varepsilon_n \mid \mathbf{z}_n \right) = 0. \tag{6}$$

This says that if we knew the true posterior means, they would be linearly related to individual characteristics $\mathbf{z}_n$ via the true vector of parameters $\boldsymbol{\delta}$, plus an idiosyncratic error ε_n. This is the relationship we seek to estimate in the second stage.

Under standard regularity conditions, the first-stage estimator is consistent and asymptotically normal:

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \Sigma_{\boldsymbol{\theta}} \right), \tag{7}$$

where $\Sigma_{\boldsymbol{\theta}}$ is estimated from the first-stage model. Because $\beta_{nk}(\boldsymbol{\theta})$ is a smooth function of $\boldsymbol{\theta}$, it can be approximated locally using a first-order Taylor expansion:

$$\hat{\beta}_{nk} \approx \beta_{nk}(\boldsymbol{\theta}) + \mathbf{G}_{nk}\left(\boldsymbol{\theta} \right) \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right) + o_p(N^{-1/2}), \tag{8a}$$

where:

$$\mathbf{G}_{nk}(\boldsymbol{\theta}) = \frac{\partial \beta_{nk}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\top}}, \tag{8b}$$

is a $1 \times P$ row vector of partial derivatives of the kth posterior mean with respect to the P hyperparameters, evaluated at the true $\boldsymbol{\theta}$. This captures how sensitive individual n's posterior mean is to changes in the population-level parameters. Because $(\hat{\beta}_{nk} - \beta_{nk}(\boldsymbol{\theta}))$ is approximately a linear function of $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$, and the latter is asymptotically normal, the former is asymptotically normal too, with variance given by the standard delta method formula:

$$\sqrt{N}\left(\hat{\beta}_{nk} - \beta_{nk}(\boldsymbol{\theta}) \right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{G}_{nk}\left(\boldsymbol{\theta} \right) \Sigma_{\boldsymbol{\theta}} \, \mathbf{G}_{nk}\left(\boldsymbol{\theta} \right)^{\top} \right). \tag{9}$$

In plain terms: the uncertainty in $\hat{\beta}_{nk}$ arising from first-stage estimation error is quantifiable, and depends on both how uncertain $\hat{\boldsymbol{\theta}}$ is and how sensitive the posterior mean is to changes in $\boldsymbol{\theta}$.

Under the further conditions that:

$$\frac{1}{N}\mathbf{Z}^{\top}\mathbf{Z} \xrightarrow{p} \mathbf{Q}_{zz}, \qquad \frac{1}{N}\sum_{n=1}^{N} z_n \mathbf{G}_{nk}\left(\boldsymbol{\theta} \right) \xrightarrow{p} \mathbf{Q}_{zg}, \tag{10}$$

with $\mathbf{Q}_{zz}$ positive definite, the asymptotic variance of the second-stage OLS estimator $\hat{\boldsymbol{\delta}}$ comprises three components:

$$\operatorname{avar}\left(\hat{\boldsymbol{\delta}} \right) = \mathbf{Q}_{zz}^{-1} \left[\Omega_{\varepsilon} + \mathbf{Q}_{zg} \Sigma_{\boldsymbol{\theta}} \mathbf{Q}_{zg}^{\top} + \mathbf{C} + \mathbf{C}^{\top} \right] \mathbf{Q}_{zz}^{-1}. \tag{11}$$

Here $\Omega_{\varepsilon} = \operatorname{plim} N^{-1} \sum_{n=1}^{N} z_n z_n^{\top} \varepsilon_n^2$ is the standard heteroskedasticity-robust variance that would be obtained if the posterior means were observed without error; $\mathbf{Q}_{zg} \Sigma_{\boldsymbol{\theta}} \mathbf{Q}_{zg}^{\top}$ is the additional variance induced by first-stage estimation error propagating through the posterior means into the second stage (i.e., precisely the term the simulation procedure below is designed to capture); and $\mathbf{C} + \mathbf{C}^{\top}$ captures the asymptotic covariance between the idiosyncratic second-stage variation, $N^{-1/2} \sum_{n=1}^{N} z_n \varepsilon_n$, and the aggregate first-stage estimation error, $\mathbf{Q}_{zg} \sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$, which in many settings will be negligible since the two sources of variation operate at different levels (one individual-specific, one aggregate), but which contributes to the true variance in general. The key implication is that naïvely running a regression on $\hat{\beta}_{nk}$ and using standard errors as if $\hat{\beta}_{nk}$ were fixed omits the second and third terms entirely, leading to underestimated standard errors and overconfident inference.

[2] Note that this is a relationship involving the posterior mean $\beta_{nk}(\boldsymbol{\theta})$ rather than the true taste parameter $\beta_{nk}^{\star}$; the two coincide only as $T_n \to \infty$, as mentioned above.

By repeatedly drawing $\boldsymbol{\theta}^{(s)} \sim \mathcal{N}(\hat{\boldsymbol{\theta}}, \Sigma_{\hat{\boldsymbol{\theta}}})$, recomputing $\hat{\beta}_{nk}^{(s)} = \beta_{nk}(\boldsymbol{\theta}^{(s)})$, and rerunning the second-stage regression each time, one is essentially simulating the distribution implied by the delta method argument above. Each s draw perturbs $\boldsymbol{\theta}$ in a way consistent with its estimated sampling distribution, and the resulting variation in $\hat{\beta}_{nk}$ across draws mimics the first-order uncertainty derived analytically above. This is precisely the approach proposed by Krinsky and Robb (1986) for propagating parameter uncertainty through nonlinear functions, adapted here to the second-stage regression context.

The derivation above assumes that the second-stage estimator is a smooth function of the posterior means, which holds for OLS and for many other estimators commonly used in practice. It is worth stressing that the foregoing discussion concerns the asymptotic properties of $\hat{\boldsymbol{\delta}}$ as a function of the estimated posterior means $\hat{\beta}_{nk}$. This is sufficient if the object of interest is inference on the relationship between posterior means and individual characteristics. However, because $\hat{\beta}_{nk}$ is the expectation derived from the posterior distribution, it does not fully reflect the uncertainty inherent in that distribution; an issue we turn to next.

Recall that the posterior mean $\hat{\beta}_{nk} = \beta_{nk}(\hat{\boldsymbol{\theta}})$ is a point estimate, specifically, the mean of the conditional distribution given the observed data and the estimated population parameters. Even if we knew $\boldsymbol{\theta}$ exactly, there would remain genuine uncertainty about the true individual-specific taste parameters $\beta_{nk}^{\star}$. All we can say is that it is drawn from some distribution $f(\cdot)$, informed by the choices observed for that individual, $\mathbf{y}_n$, the attributes of the alternatives they faced, $\mathbf{X}_n$, and the parameters describing the population-level distribution of preferences, $\boldsymbol{\theta}$:

$$\boldsymbol{\beta}_{nk} \mid \mathbf{y}_n, \mathbf{X}_n, \boldsymbol{\theta} \sim f(\cdot). \tag{12}$$

The more choices individual n made (larger T_n), and the greater the variation in the attributes they faced, the more informative $\mathbf{y}_n$ and $\mathbf{X}_n$ become, and the tighter this distribution concentrates around the true $\beta_{nk}^{\star}$ (Sarrias, 2020). In the limit, as $T_n \to \infty$, the posterior collapses to a point mass at $\beta_{nk}^{\star}$. Separately, as $N \to \infty$, consistency of the first-stage estimator ensures $\hat{\boldsymbol{\theta}} \to \boldsymbol{\theta}$, eliminating the additional uncertainty introduced by plug-in estimation of the hyperparameters. In addition to these asymptotic properties, it should be recognised that posterior means from conditional distributions tend to "shrink" toward the population mean, especially when individuals provide limited information (e.g., few choices or little variation). This occurs because the conditional means are influenced by the estimated population parameters, which reflect averages across individuals. As a result, the distribution of conditional means is typically narrower than the unconditional distribution (e.g., Revelt and Train, 2000; Hess, 2010; Sarrias and Daziano, 2018). This narrowing arises because conditional means do not capture the variance around themselves, which contributes to the full unconditional variance (Daly et al., 2012). Such shrinkage can introduce attenuation bias in second-stage regressions that treat these means as true individual parameters, another reason why its important to consider the full posterior distribution rather than just the point estimates when conducting second-stage analyses, especially when the number of choice observations per individual is small, and the degree of shrinkage is likely to be substantial.

To formalise this second source of uncertainty, define the conditional variance vector for individual n as:

$$\Lambda_n(\boldsymbol{\theta}) \equiv \mathrm{Var}_{\boldsymbol{\theta}}\left(\boldsymbol{\beta}_n \mid \mathbf{y}_n, \mathbf{X}_n\right) = \mathbb{E}_{\boldsymbol{\theta}}\left[\left(\boldsymbol{\beta}_n - \beta_n(\boldsymbol{\theta})\right)\left(\boldsymbol{\beta}_n - \beta_n(\boldsymbol{\theta})\right)^{\top} \mid \mathbf{y}_n, \mathbf{X}_n\right]. \tag{13}$$

This matrix captures the spread of the posterior distribution around its mean, reflecting the irreducible uncertainty that remains even if the population hyperparameters $\boldsymbol{\theta}$ were known exactly. When individual n provides many choice observations (large T_n) with sufficient variation in the choice attributes, the likelihood dominates the prior and $\Lambda_n(\boldsymbol{\theta}) \to \mathbf{0}$ as $T_n \to \infty$, meaning the posterior collapses and the plug-in posterior mean $\hat{\beta}_{nk}$ converges to the true individual-specific parameter $\beta_{nk}^{\star}$. However, in finite samples where T_n is small or fixed (as is typically the case in practice), $\Lambda_n(\boldsymbol{\theta})$ remains strictly positive. Because the point summary $\hat{\beta}_{nk}$ fails to capture this spread, treating it as the true parameter ignores a second, distinct source of uncertainty, that is entirely separate from the first-stage hyperparameter uncertainty discussed above, which can lead to further underestimation of standard errors in subsequent second-stage analyses.

This gives rise to a decomposition of the total discrepancy between the plug-in posterior mean and the true individual taste parameter into two distinct components:

$$\left(\hat{\beta}_{nk} - \beta_{nk}^{\star}\right) = \underbrace{\left[\hat{\beta}_{nk} - \beta_{nk}(\boldsymbol{\theta})\right]}_{\substack{\text{hyperparameter}\\ \text{uncertainty}}} + \underbrace{\left[\beta_{nk}(\boldsymbol{\theta}) - \beta_{nk}^{\star}\right]}_{\substack{\text{posterior}\\ \text{shrinkage}}}. \tag{14}$$

The first term (hyperparameter uncertainty) arises from sampling variability in the first-stage parameter estimates and is precisely what the simulation procedure outlined above addresses, as formalised by the delta method argument above. The second term (posterior shrinkage) reflects the irreducible uncertainty about $\boldsymbol{\beta}_n^{\star}$ that remains even at the true $\boldsymbol{\theta}$. It arises because $\hat{\beta}_{nk}$ is only a point summary of the posterior distribution, which has positive variance $\left[\Lambda_n(\boldsymbol{\theta})\right]_{kk}$ in any finite sample. More generally, when multiple posterior means are used jointly in a second-stage analysis, the full posterior variance matrix $\Lambda_n(\boldsymbol{\theta})$ is relevant, as its off-diagonal elements $\left[\Lambda_n(\boldsymbol{\theta})\right]_{kj}$ capture posterior covariance across attributes k and j.

To account for both sources of uncertainty simultaneously, one can draw directly from the full posterior distribution $\hat{F}_{nk}\left(\beta \mid \hat{\boldsymbol{\theta}}\right)$ rather than simply recomputing the posterior mean at perturbed hyperparameter values. Specifically, for each simulation draw s:

1. Draw $\boldsymbol{\theta}^{(s)} \sim \mathcal{N}(\hat{\boldsymbol{\theta}}, \Sigma_{\hat{\boldsymbol{\theta}}})$, propagating hyperparameter uncertainty as before.
2. For each individual n, obtain a draw $\tilde{\beta}_{nk}^{(s)}$ from the posterior distribution recomputed at the perturbed hyperparameters:

$$\tilde{\beta}_{nk}^{(s)} \sim \hat{F}_{nk}\left(\beta \mid \boldsymbol{\theta}^{(s)}\right), \quad \text{where} \quad \hat{F}_{nk}\left(\beta \mid \boldsymbol{\theta}^{(s)}\right) \equiv \left\{\left(\beta_{nk1}^{(s)}, \omega_{n1}^{(s)}\right), \left(\beta_{nk2}^{(s)}, \omega_{n2}^{(s)}\right), \ldots, \left(\beta_{nkR}^{(s)}, \omega_{nR}^{(s)}\right)\right\}, \tag{15}$$

where the support points $\beta_{nkr}^{(s)}$ and weights $\omega_{nr}^{(s)}$ are recomputed using $\boldsymbol{\theta}^{(s)}$ in place of $\hat{\boldsymbol{\theta}}$.

3. Rerun the second-stage regression using $\tilde{\beta}_{nk}^{(s)}$ in place of $\hat{\beta}_{nk}$.

This extended procedure propagates both sources of uncertainty into the second-stage estimates. The asymptotic variance of the resulting estimator $\hat{\boldsymbol{\delta}}$ now comprises an additional term relative to the expression derived above:

$$\text{avar}\left(\hat{\boldsymbol{\delta}}\right) = \mathbf{Q}_{zz}^{-1}\left[\Omega_{\epsilon} + \mathbf{Q}_{zg}\Sigma_{\boldsymbol{\theta}}\mathbf{Q}_{zg}^{\top} + \mathbf{Q}_{z\lambda} + \mathbf{C} + \mathbf{C}^{\top}\right]\mathbf{Q}_{zz}^{-1}, \tag{16}$$

where $\mathbf{Q}_{z\lambda} = \text{plim}\ N^{-1}\sum_{n=1}^{N} z_n \Lambda_n(\boldsymbol{\theta}) z_n^{\top}$ captures the contribution of the posterior spread to the asymptotic variance of the second-stage estimator. Intuitively, when the posterior distributions are diffuse (e.g., because individuals provide limited information) the additional noise introduced by drawing from those distributions inflates the variance of the second-stage estimates beyond what hyperparameter uncertainty alone would imply.

It is important to recognise what each simulation procedure achieves and what it does not. The first procedure, which draws only from the sampling distribution of $\hat{\boldsymbol{\theta}}$ and recomputes posterior means, correctly captures the additional variance term $\mathbf{Q}_{zg}\Sigma_{\boldsymbol{\theta}}\mathbf{Q}_{zg}^{\top}$ term induced by first-stage estimation error, but leaves the posterior spread term $\mathbf{Q}_{z\lambda}$ unaddressed. The extended procedure, which additionally draws from the conditional distribution $\hat{F}_{nk}(\beta \mid \boldsymbol{\theta}^{(s)})$, captures both. The choice between them depends on the inferential objective. If the goal is inference on a relationship involving the posterior mean $\beta_{nk}(\boldsymbol{\theta})$ (i.e., treating the posterior mean itself as the object of interest) the first procedure suffices. If the goal is inference on a relationship involving the true individual-specific taste parameter $\beta_{nk}^{\star}$, the extended procedure is required. Even then, a sufficiently large within-individual sample size T_n is needed to ensure that the posterior variance $\left[\Lambda_n(\boldsymbol{\theta})\right]_{kk}$ is small enough for the conditional distribution to be tightly concentrated around $\beta_{nk}^{\star}$, so that draws from it serve as reliable proxies for the true parameter. In practice, when T_n is small or fixed, neither procedure fully resolves the target mismatch between $\beta_{nk}(\boldsymbol{\theta})$ and $\beta_{nk}^{\star}$, and this limitation should be acknowledged when interpreting second-stage results.

2.3. Simulation procedures

The theoretical exposition above identifies two distinct sources of uncertainty that affect second-stage inference: hyperparameter uncertainty, captured by the term $\mathbf{Q}_{zg}\Sigma_{\boldsymbol{\theta}}\mathbf{Q}_{zg}^{\top}$, and posterior spread, captured by $\mathbf{Q}_{z\lambda}$. To illustrate the practical implications of addressing neither, one, or both of these sources, we construct four alternative empirical sampling distributions of the second-stage estimator $\hat{\boldsymbol{\delta}}$, corresponding to the four possible combinations of accounting or not accounting for each source of uncertainty.

1. ***Ignoring both sources of uncertainty:*** The second-stage regression is estimated using only the plug-in posterior means $\hat{\beta}_{nk} = \beta_{nk}(\hat{\boldsymbol{\theta}})$ as the dependent variable, treating them as fixed and known. This produces the naïve estimator whose standard errors omit both the $\mathbf{Q}_{zg}\Sigma_{\boldsymbol{\theta}}\mathbf{Q}_{zg}^{\top}$ and $\mathbf{Q}_{z\lambda}$ terms, and will in general understate the true sampling variability of $\hat{\boldsymbol{\delta}}$.
2. ***Hyperparameter uncertainty only:*** The second-stage analysis is replicated across S draws of the first-stage hyperparameter vector from its estimated asymptotic sampling distribution:

$$\boldsymbol{\theta}^{(s)} \sim \mathcal{N}\left(\hat{\boldsymbol{\theta}}, \Sigma_{\hat{\boldsymbol{\theta}}}\right), \quad s = 1, \ldots, S, \tag{17}$$

   with posterior means $\beta_{nk}(\boldsymbol{\theta}^{(s)})$ recomputed at each draw and used as the dependent variable in the second-stage regression. This propagates hyperparameter uncertainty into the second-stage estimates, capturing the $\mathbf{Q}_{zg}\Sigma_{\boldsymbol{\theta}}\mathbf{Q}_{zg}^{\top}$ term, but leaves posterior spread $\mathbf{Q}_{z\lambda}$ unaddressed.[3]
3. ***Posterior spread only:*** The second-stage analysis is replicated using draws $\tilde{\beta}_{nk}^{(s)}$ taken directly from each individual's estimated posterior distribution $\hat{F}_{nk}(\beta \mid \hat{\boldsymbol{\theta}})$, with the hyperparameters held fixed at $\hat{\boldsymbol{\theta}}$:

$$\tilde{\beta}_{nk}^{(s)} \sim \hat{F}_{nk}\left(\beta \mid \hat{\boldsymbol{\theta}}\right) \quad \text{where} \quad \hat{F}_{nk}\left(\beta \mid \hat{\boldsymbol{\theta}}\right) \equiv \left\{\left(\beta_{nk1}, \omega_{n1}\right), \left(\beta_{nk2}, \omega_{n2}\right), \ldots, \left(\beta_{nkR}, \omega_{nR}\right)\right\}. \tag{18}$$

   This accounts for uncertainty in the conditional distribution (capturing the $\mathbf{Q}_{z\lambda}$ term) but does not incorporate sampling error in $\hat{\boldsymbol{\theta}}$, leaving the $\mathbf{Q}_{zg}\Sigma_{\boldsymbol{\theta}}\mathbf{Q}_{zg}^{\top}$ term unaddressed.
4. ***Both sources of uncertainty:*** The second-stage analysis is replicated accounting for both sources simultaneously, following the extended procedure outlined above. For each draw $s = 1, \ldots, S$:

   i. Draw $\boldsymbol{\theta}^{(s)} \sim \mathcal{N}(\hat{\boldsymbol{\theta}}, \Sigma_{\hat{\boldsymbol{\theta}}})$, propagating hyperparameter uncertainty.
   ii. Recompute the posterior distribution $\hat{F}_{nk}(\beta \mid \boldsymbol{\theta}^{(s)})$ for each individual n, updating both the support points $\beta_{nkr}^{(s)}$ and the conditional weights $\omega_{nr}^{(s)}$ using $\boldsymbol{\theta}^{(s)}$.
   iii. Obtain a draw $\tilde{\beta}_{nk}^{(s)} \sim \hat{F}_{nk}(\beta \mid \boldsymbol{\theta}^{(s)})$, propagating posterior spread.
   iv. Rerun the second-stage regression using $\tilde{\beta}_{nk}^{(s)}$ as the dependent variable and retain $\hat{\boldsymbol{\delta}}^{(s)}$.

This comprehensive procedure captures both $\mathbf{Q}_{zg}\Sigma_{\boldsymbol{\theta}}\mathbf{Q}_{zg}^{\top}$ and $\mathbf{Q}_{z\lambda}$, yielding the most accurate empirical sampling distribution for $\hat{\boldsymbol{\delta}}$. The distribution of $\hat{\boldsymbol{\delta}}^{(s)}$ across the S replications provides simulation-consistent standard errors and confidence intervals for the second-stage estimator.

[3] This is akin to the resampling approach of Chappell et al. (2022), who develop a simulation-based procedure to obtain valid asymptotic standard errors for individual-level posterior parameter estimates recovered from a random parameters ordered probit model. Our procedure differs in that the focus is on propagating first-stage uncertainty into a second-stage regression analysis.

2.4. Diagnosing consistency and efficiency via bootstrapping

The theoretical results above establish that the naïve second-stage estimator (which treats $\hat{\beta}_{nk}$ as fixed and known) omits the variance terms $\mathbf{Q}_{zg}\Sigma_{\boldsymbol{\theta}}\mathbf{Q}_{zg}^{\top}$ and $\mathbf{Q}_{z\lambda}$, leading to understated standard errors. However, omitting these terms need not compromise the consistency of $\hat{\boldsymbol{\delta}}$. In the OLS setting considered here, where $\hat{\beta}_{nk}$ enters as the dependent variable, measurement error in $\hat{\beta}_{nk}$ is absorbed into the regression residual rather than the regressors $\mathbf{z}_n$. Provided $\mathbb{E}(\epsilon_n \mid \mathbf{z}_n) = 0$ holds and T_n is sufficiently large that $\beta_{nk}(\boldsymbol{\theta}) \approx \beta_{nk}^{\star}$, OLS remains consistent for $\boldsymbol{\delta}$, albeit inefficient. As a consequence, the simulation-based estimates $\hat{\boldsymbol{\delta}}^{(s)}$ will be centred on $\hat{\boldsymbol{\delta}}$ across replications, confirming consistency rather than correcting for any bias.

This property does not, however, generalise to all second-stage settings. When posterior means enter as generated regressors on the right-hand side of the second-stage model, measurement error is no longer absorbed into the residual but instead contaminates the regressor, inducing attenuation bias and rendering the naïve estimator inconsistent (see, e.g., Pagan, 1984). Similarly, in nonlinear second-stage models the relationship between the generated variable and the structural parameter of interest can be sufficiently complex that the naïve estimator does not converge to the correct quantity. In both cases, the simulation procedure reproduces rather than corrects the underlying bias: it recovers the empirical sampling distribution of a biased estimator, not of a consistent one. Researchers working in these settings should therefore exercise considerable caution, and may need to adopt bias-correction methods or instrumental variable approaches before applying the simulation procedure.

To assess consistency and efficiency empirically, we propose the following bootstrap procedure applied to the S second-stage estimates $\{\hat{\boldsymbol{\delta}}^{(s)}\}_{s=1}^{S}$ obtained under the three simulation procedures.[4]

Consistency

i. *Compare point estimates:* Compute the naïve plug-in estimate $\hat{\boldsymbol{\delta}}$ to the mean of $\{\hat{\boldsymbol{\delta}}^{(s)}\}_{s=1}^{S}$. Agreement between the two provides empirical evidence of consistency.
ii. *Bootstrap confidence interval:* Resample $\{\hat{\boldsymbol{\delta}}^{(s)}\}_{s=1}^{S}$ with replacement, computing the mean of each resample, and take the 2.5th and 97.5th percentiles across resamples to form a 95% bootstrap confidence interval for $\boldsymbol{\delta}$.
iii. *Assess consistency:* If $\hat{\boldsymbol{\delta}}$ lies within this interval, the naïve estimator is consistent (i.e., it recovers the correct parameter values despite ignoring the two additional variance terms). If it lies outside, the omitted terms affect not just efficiency but the point estimate itself, indicating inconsistency.

Efficiency

i. *Bootstrap the confidence intervals bounds:* For each bootstrap sample, compute a 95% confidence interval for $\boldsymbol{\delta}$. Across all bootstrap samples, compute the 2.5th and 97.5th percentiles of both the lower and upper bounds separately, yielding a confidence interval around each boundary of the bootstrap interval (i.e., an interval for the interval).
ii. *Compare with naïve inference:* Construct the classical confidence interval based on the naïve estimator:

$$\hat{\boldsymbol{\delta}} \pm t_{N-K,0.025} \cdot \mathrm{SE}\left(\hat{\boldsymbol{\delta}}\right), \tag{19}$$

where $t_{N-K,0.025}$ is the t-critical value with $N-K$ degrees of freedom (relating to the second stage model) and $\mathrm{SE}(\hat{\boldsymbol{\delta}})$ is the naïve standard error that ignores both $\mathbf{Q}_{zg}\Sigma_{\boldsymbol{\theta}}\mathbf{Q}_{zg}^{\top}$ and $\mathbf{Q}_{z\lambda}$.
iii. *Interpret the comparison:* If the naïve confidence interval lies entirely within the bootstrapped bounds (i.e., the bootstrapped lower limit falls below and the bootstrapped upper limit extends above the naïve interval) this indicates that the naïve estimator is consistent but inefficient: it is centred correctly but understates the true sampling variability of $\hat{\boldsymbol{\delta}}$ by omitting the additional variance terms identified previously.

3. Demonstration

3.1. Data and models

To illustrate our approach, we use stated choice experiment data from Norway designed to elicit marginal willingness to pay (WTP) to prevent the spread of an invasive crab species. Further details on the data and study results can be found in Sandorf et al. (2025). The discrete choice experiment consisted of 40 binary choice tasks, divided into five blocks. The sample comprised 1496 respondents, each completing eight choice tasks, yielding a total of 11,968 choice observations. Each alternative is described by four attributes: spread (presence of the invasive crab, coded as four dummy variables), species remaining in the area with the crab (continuous percentage of species remaining.), environmental quality (coded using four dummy variables), and tax (which is the yearly nominal increase in taxes for a period of 15 years, entered as a continuous variable).

A standard conditional logit model using these attributes (plus a status-quo constant) includes 11 parameters. To focus on unobserved preference heterogeneity, rather than interacting attributes with respondent characteristics, we estimate a mixed logit (random-parameters logit) model. All 11 parameters are treated as random and normally distributed, except the cost (tax) coefficient,

[4] This is a non-parametric bootstrap applied to the set of second-stage estimates obtained across the S simulation draws. It does not involve resampling individuals or re-estimating the first-stage model, and is therefore computationally straightforward.

which follows a negative log-normal distribution. Allowing full correlation via a Cholesky decomposition yields 77 parameters. The model is estimated using 2500 Modified Latin Hypercube Sampling draws (Hess et al., 2006). Because all inference procedures occur post-estimation, the model needs to be estimated only once, keeping computation manageable. In the second-stage, we use the individual-specific posterior distributions of marginal WTP implied by the model. We focus on WTP for the "species remains" attribute, defined as the negative ratio of its random coefficient, β^b, to the random cost coefficient, β^c. Thus, the posterior distribution of marginal WTP for individual n is:

$$\hat{F}_{nk}\left(\hat{wtp}_n^b \mid \hat{\boldsymbol{\theta}}\right) \equiv \left\{ \left(-\frac{\beta_{n1}^b}{\beta_{n1}^c}, \omega_{n1} \right), \left(-\frac{\beta_{n2}^b}{\beta_{n2}^c}, \omega_{n2} \right), \ldots, \left(-\frac{\beta_{nR}^b}{\beta_{nR}^c}, \omega_{nR} \right) \right\}. \tag{20}$$

We treat these individual-level WTP estimates as the dependent variable in an OLS regression:

$$\hat{wtp}_n^b = \hat{\delta}_0 + \hat{\delta}_1 m_n + \hat{\delta}_2 a_n, \tag{21}$$

where $m_n = 1$ if respondent n is male (0 otherwise), and a_n is age in years. The key parameters of interest are $\hat{\delta}_1$ and $\hat{\delta}_2$, and we test the null hypotheses $H_0 : \delta_1 = 0$ and $H_0 : \delta_2 = 0$ using two-tailed tests.

We begin with the naïve approach: regressing posterior means of individual-specific marginal WTPs on gender and age. This yields point estimates and a variance–covariance matrix from which standard errors and 95% confidence intervals can be constructed under standard OLS assumptions. To evaluate the validity of the naïve inference, we implement the three simulation-based procedures outlined earlier, which account for uncertainty in both the conditional marginal WTP distributions and the first-stage parameter estimates. The core R function for simulating the conditional distributions from `apollo` (Hess and Palma, 2025) output is provided in Appendix A.

3.2. Findings

Complete estimation results for the mixed logit model, including plots illustrating the resulting heterogeneity in marginal WTP and the corresponding OLS regression output, are provided in Appendix B. For brevity, we focus here on the empirical sampling distributions of the second-stage estimates δ_1 and δ_2, presented in Fig. 1(a) and Fig. 1(b), respectively. For the naïve approach, the sampling distribution is generated from a t-distribution with mean $\hat{\delta}$, standard error SE($\hat{\delta}$), and 1494 degrees of freedom. The three simulation-based empirical sampling distributions are each constructed from 5000 replications of the second-stage regression under Procedures 2–4, and displayed as kernel density estimates. Vertical lines mark the mean and 95% confidence interval bounds (2.5th and 97.5th percentiles), with shaded envelopes indicating the 95% confidence region. Horizontal bars beneath each panel report the widths of the 90% and 95% confidence intervals to facilitate comparison of estimator efficiency across procedures.

The results are instructive. Under the naïve approach, both $\hat{\delta}_1$ and $\hat{\delta}_2$ are statistically significant, suggesting meaningful demographic effects on marginal WTP. Accounting for hyperparameter uncertainty alone (Procedure∼2) yields near-identical conclusions and, if anything, a slightly tighter sampling distribution. This is a reminder that the naïve standard error is not simply too small, but a biased estimate of the wrong quantity: the naïve OLS residual conflates true idiosyncratic error with measurement error, inflating the residual variance, while the simulation propagates only hyperparameter uncertainty, which is genuinely small when the first-stage model is estimated precisely.[5] Introducing posterior spread (Procedure 3) substantially widens the sampling distributions, however, such that neither demographic effect retains significance; revealing a Type I error under the naïve approach. Accounting for both sources simultaneously (Procedure 4) widens the distributions further, reinforcing this conclusion.

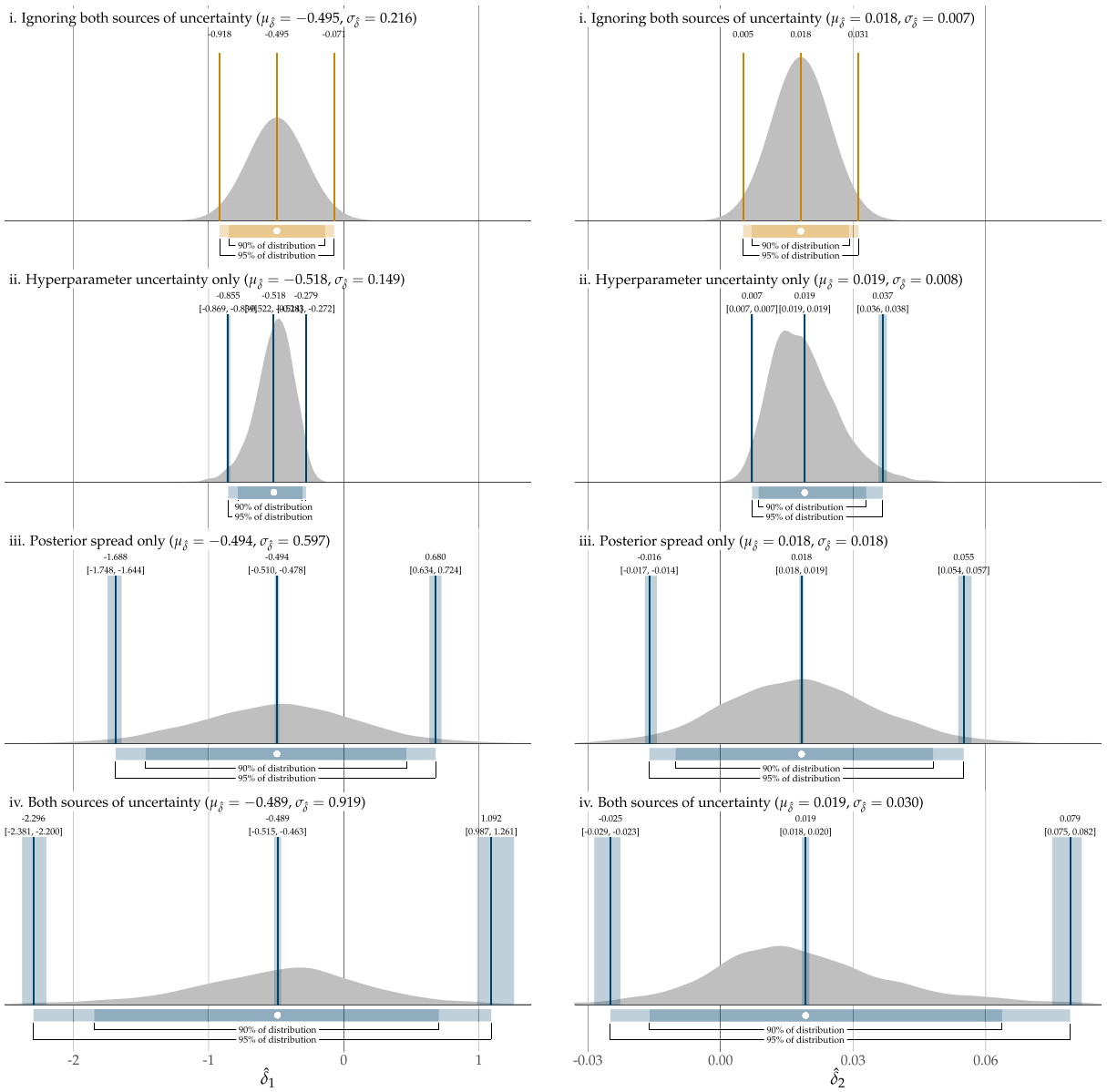
Despite these differences in precision, all four approaches yield closely aligned point estimates, with the naïve $\hat{\delta}$ generally falling within the simulation-based confidence intervals. This provides empirical support for the consistency of the naïve estimator. The systematically narrower naïve confidence intervals confirm, however, that the issue is one of efficiency rather than consistency: the naïve approach recovers the correct parameter values but understates their true sampling variability, leading to overconfident inference.

4. Conclusion

This paper addresses a well-known but frequently overlooked problem in applied discrete choice modelling: the proper treatment of uncertainty in second-stage analyses that use individual-level posterior means from mixed logit models. Although the sampling variability of first-stage estimates and the spread of conditional distributions have been noted (e.g., Revelt and Train, 2000; Train, 2003; Hess, 2010; Sarrias and Daziano, 2018; Chappell et al., 2022), their combined impact on downstream inference has received little systematic attention. Treating plug-in posterior means $\hat{\beta}_{nk}$ as fixed and known omits two distinct variance components, hyperparameter uncertainty and posterior spread, leading to understated standard errors and overconfident inference.

Grounding both sources of uncertainty in a delta method framework, we propose a Krinsky–Robb type simulation procedure (Krinsky and Robb, 1986) to recover the empirical sampling distribution of second-stage estimators. The empirical case study confirms that naïve estimators are consistent but inefficient: ignoring either source of uncertainty produces standard errors that

[5] There is no theoretical guarantee that the simulation-based distribution must be wider than the naïve one. A negative cross-covariance between first-stage estimation error and second-stage idiosyncratic noise ($\mathbf{C} + \mathbf{C}^\top < 0$) can further reduce the total variance below Ω_ε alone. The direction of the difference depends on the specific data and model.



(a) Male (δ_1) (b) Age (δ_2)

Fig. 1. Sampling distributions of the second-stage estimates.

are materially understated. Although the exposition focuses on OLS, the procedure extends to any second-stage estimator that is a smooth function of the posterior means and for which repeated estimation is feasible (including a wide range of estimators commonly used in practice, such as *t*-tests, ANOVA, and spatial models) with the same delta method justification carrying over directly. In the multivariate case, where posterior means for several attributes are used simultaneously (e.g., Campbell, 2007), the procedure must additionally account for the full posterior covariance matrix $\Lambda_n(\boldsymbol{\theta})$. Formal proofs for each estimator class and a thorough treatment of the multivariate case remain important avenues for future work.

Two limitations deserve emphasis. First, when T_n is small, neither procedure fully resolves the target mismatch between $\beta_{nk}(\boldsymbol{\theta})$ and $\beta^{\star}_{nk}$: improving inference about posterior means is not the same as recovering relationships involving true taste parameters. Second, once systematic preference variation is identified in a second-stage analysis, the natural response is to incorporate those covariates directly into the mixed logit specification. Re-estimating the model with these relationships embedded allows the researcher to capture the directional effects of observed heterogeneity on preferences within a coherent structural framework, rather than treating them as post-hoc observations. Where re-estimation is not feasible, practitioners should be especially cautious about

using second-stage relationships to predict preferences for unsampled individuals or populations: such predictions are conditional on realised choices and subject to posterior shrinkage, and do not carry the same inferential standing as structurally identified parameters. The procedures developed in this paper ensure that the two sources of uncertainty are correctly propagated, but correct propagation is a necessary condition for valid inference, not a sufficient one: whether the recovered relationships are structurally meaningful remains a substantive question that no amount of simulation can resolve.

CRediT authorship contribution statement

Danny Campbell: Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Erlend Dancke Sandorf:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Tobias Börger:** Writing – review & editing, Methodology, Conceptualization. **Thijs Dekker:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Danny Campbell acknowledges the research funding that supported part of his research time, specifically under the CAVEAT project (AH/Y000528/1), jointly funded by the UK Arts and Humanities Research Council and the UK Department for Digital, Culture, Media and Sport.

Appendix A. R code for conditional draws from apollo model estimates

This appendix provides the primary R code used to generate the results presented in the main text, centred on the `conditional_draw` function. The function is designed for use with model output from the `apollo` package (Hess and Palma, 2025) and can be executed in any R environment where the package is installed and loaded. Its function arguments are:

- `model`: the fitted `apollo` model object, including estimated parameters and their (robust) variance–covariance matrix.
- `apollo_probabilities`: the user-defined function specifying the model's choice probabilities.
- `apollo_inputs`: the structured list created by `apollo_validateInputs()`, containing all data, parameter names, and settings required by `apollo_probabilities`.
- `use_varcov`: logical; if TRUE (default), draws incorporate uncertainty in the first-stage parameter estimates.
- `return_output`: logical; if TRUE (default), the function returns a list containing (i) a dataframe of conditional draws, (ii) the unconditional draws, and (iii) the corresponding weights. If FALSE, only the conditional-draw dataframe is returned.

```
# ------------------------------------------------------------------------------
# conditional_draw()
# ------------------------------------------------------------------------------
# Generates one draw from the individual-level conditional distribution implied
# by an apollo model, optionally incorporating uncertainty in the first-stage
# parameter estimates. Supports random parameters and latent class logit models.
#
# Args:
#   model              : apollo model object returned by apollo_estimate()
#   apollo_probabilities : Probability function passed to apollo_estimate()
#   apollo_inputs      : Model input list
#   use_varcov         : If TRUE, simulate parameters using the robust VC matrix
#   return_output      : If TRUE, return weights & unconditional dists as well
#
# Returns:
#   A data frame of one simulated conditional draw for each individual and
#   random parameter (or LC class), optionally with weights and unconditionals.
# ------------------------------------------------------------------------------

conditional_draw <- function(model, apollo_probabilities, apollo_inputs,
                             use_varcov = TRUE, return_output = TRUE) {

  # Copy model so simulated coefficients do not overwrite original
  model_draw <- model
```

```
# ------------------------------------------------------------------------------
# 1. Simulate new parameter vector from asymptotic normal distribution
# ------------------------------------------------------------------------------
if (use_varcov) {
  coef_names <- setdiff(names(model$estimate), model$apollo_fixed)

  # Cholesky of robust variance-covariance matrix
  L <- chol(model$robvarcov)

  # Draw standard normal vector
  z <- matrix(rnorm(length(coef_names)), nrow = 1)

  # Transform to simulated coefficient draw
  sim_draw <- z %*% L + matrix(model$estimate[coef_names], nrow = 1)

  # Assign simulated coefficients
  model_draw$estimate[coef_names] <- sim_draw[1, ]
}

# Unpack objects from apollo_inputs
silent         <- isTRUE(apollo_inputs$silent)
apollo_beta    <- model_draw$estimate
apollo_fixed   <- model$apollo_fixed
apollo_control <- apollo_inputs[["apollo_control"]]
database       <- apollo_inputs[["database"]]
draws          <- apollo_inputs[["draws"]]
apollo_randCoeff <- apollo_inputs[["apollo_randCoeff"]]
apollo_draws   <- apollo_inputs[["apollo_draws"]]
apollo_lcPars  <- apollo_inputs[["apollo_lcPars"]]

continuous   <- isTRUE(apollo_control$mixing)
latentClass  <- is.function(apollo_lcPars)

if (is.null(apollo_control$HB)) {
  stop("Not implemented for HB models.")
}

# ------------------------------------------------------------------------------
# 2. Compute unconditional distributions & conditional weights
# ------------------------------------------------------------------------------

if (continuous) {
  # Mixed logit: unconditional distribution of random parameters
  unconds_draw <- apollo_mixUnconditionals(
    model_draw, apollo_probabilities, apollo_inputs
  )

  # Conditional probabilities for each draw
  P <- apollo_probabilities(
    model_draw$estimate,
    apollo_inputs,
    functionality = "conditionals"
  )

  # Conditional weights (posterior probabilities)
  w <- P / rowSums(P)
}

if (latentClass) {
  # Latent class: conditional weights over classes
  w <- apollo_lcConditionals(model_draw, apollo_probabilities,
```

```
                                apollo_inputs)[, -1]

    # Unconditional distributions for each class
    unconds_draw <- apollo_lcUnconditionals(model_draw, apollo_probabilities,
                                            apollo_inputs)

    # Remove mixing weights (not needed here)
    unconds_draw$pi_values <- NULL

    # Expand to individual x class structure
    n_indivs <- model$nIndivs
    n_col <- ncol(w)

    unconds_draw <- lapply(unconds_draw, function(x) {
      mat <- matrix(unlist(x), ncol = n_col, byrow = TRUE)
      mat[rep(1L, n_indivs), , drop = FALSE]
    })
  }

  # ------------------------------------------------------------------------------
  # 3. Generate one conditional draw per individual
  # ------------------------------------------------------------------------------

  n_row <- nrow(w)
  n_var <- length(unconds_draw)

  conds_draw <- matrix(NA_real_, nrow = n_row, ncol = n_var)
  colnames(conds_draw) <- names(unconds_draw)

  # Draw one component/class per individual using conditional probabilities
  rand_indices <- apply(w, 1, function(p) sample.int(length(p), 1L, prob = p))

  seq_n <- seq_len(n_row)
  for (i in seq_along(unconds_draw)) {
    conds_draw[, i] <- unconds_draw[[i]][cbind(seq_n, rand_indices)]
  }

  # ------------------------------------------------------------------------------
  # 4. Return formatted output
  # ------------------------------------------------------------------------------
  if (return_output) {
    return(list(
      cond_draws = as.data.frame(conds_draw),
      unconds    = unconds_draw,
      weights    = w
    ))
  } else {
    return(as.data.frame(conds_draw))
  }
}
```

Appendix B. Model results

Estimation results from the random parameters logit model are reported in Table B.1.

In Fig. B.1, we show histograms of the marginal willingness to pay (WTP) for the "species remains" attribute (in NOK per person per year). The top panel presents the unconditional WTP distribution, reflecting overall preference heterogeneity in the sample. The four lower panels display conditional mean WTP distributions by gender and age: males ≤50, males >50, females ≤50, and females >50. These plots illustrate subgroup differences in individual-level WTP implied by the model. The extent to which these differences are systematic is assessed formally in the second-stage OLS analysis.

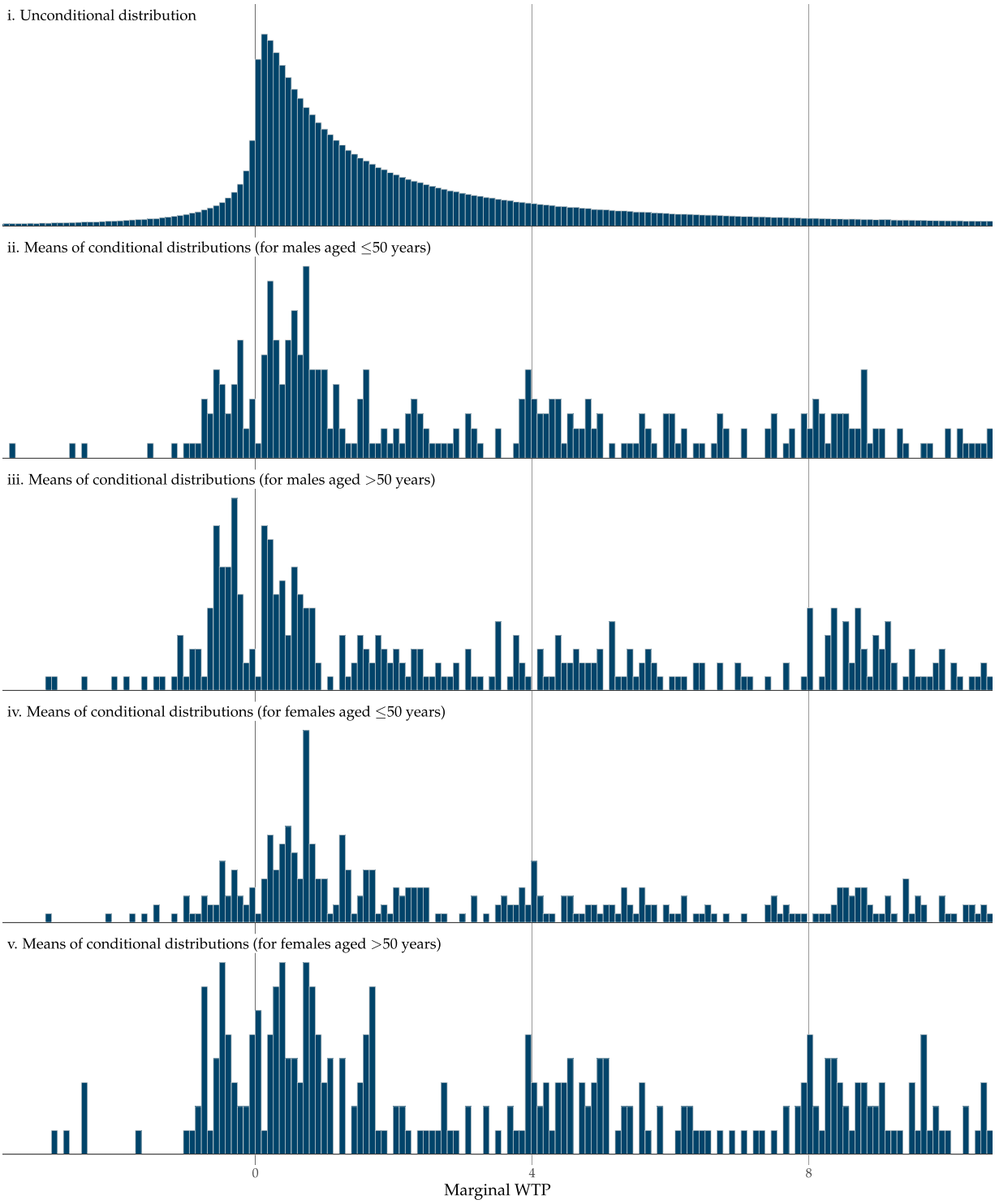


Fig. B.1. Marginal willingness to pay (WTP) for the "species remains" attribute (in NOK per person per year)

Table B.1
Mixed logit model results.

Parameter	Estimate	Parameter	Estimate	Parameter	Estimate	Parameter	Estimate
mu_sq	2.713 (41.364)	mu_sp100	−0.475*** (0.100)	mu_sp200	−1.001*** (0.164)	mu_sp300	−1.652*** (0.227)
mu_sp400	−1.921*** (0.291)	mu_spc	3.116*** (0.265)	mu_evql2	1.195*** (0.140)	mu_evql3	2.497*** (0.206)
mu_evql4	3.534*** (0.293)	mu_evql5	3.995*** (0.332)	mu_tax	0.551*** (0.087)	ch_sq	−2.806 (41.376)
ch_sq_sp100	0.312 (0.220)	ch_sp100	−0.894*** (0.176)	ch_sq_sp200	1.030*** (0.339)	ch_sp100_sp200	−1.167*** (0.366)
ch_sp200	0.283 (0.261)	ch_sq_sp300	1.787*** (0.320)	ch_sp100_sp300	−1.327** (0.552)	ch_sp200_sp300	−0.259 (0.349)
ch_sp300	−0.332 (0.376)	ch_sq_sp400	1.856*** (0.395)	ch_sp100_sp400	−1.402** (0.670)	ch_sp200_sp400	−0.144 (0.523)
ch_sp300_sp400	−0.696 (0.554)	ch_sp400	0.499* (0.280)	ch_sq_spc	−0.680 (0.423)	ch_sp100_spc	0.794** (0.402)
ch_sp200_spc	−0.388 (0.812)	ch_sp300_spc	−0.691 (0.549)	ch_sp400_spc	−1.287*** (0.446)	ch_spc	1.860*** (0.405)
ch_sq_evql2	−0.195 (0.364)	ch_sp100_evql2	0.204 (0.422)	ch_sp200_evql2	−0.492 (0.486)	ch_sp300_evql2	0.202 (0.548)
ch_sp400_evql2	−0.114 (0.525)	ch_spc_evql2	1.121*** (0.349)	ch_evql2	0.859** (0.389)	ch_sq_evql3	−1.207*** (0.457)
ch_sp100_evql3	−0.171 (0.338)	ch_sp200_evql3	−1.143*** (0.429)	ch_sp300_evql3	0.084 (0.430)	ch_sp400_evql3	−0.370 (0.393)
ch_spc_evql3	1.605*** (0.399)	ch_evql2_evql3	0.500 (0.587)	ch_evql3	−0.005 (0.331)	ch_sq_evql4	−1.562*** (0.575)
ch_sp100_evql4	−0.112 (0.405)	ch_sp200_evql4	−0.629 (0.525)	ch_sp300_evql4	0.383 (0.529)	ch_sp400_evql4	−0.442 (0.593)
ch_spc_evql4	2.432*** (0.401)	ch_evql2_evql4	0.212 (0.871)	ch_evql3_evql4	−0.229 (0.309)	ch_evql4	−0.095 (0.242)
ch_sq_evql5	−2.051*** (0.702)	ch_sp100_evql5	−0.240 (0.534)	ch_sp200_evql5	−0.672 (0.618)	ch_sp300_evql5	0.831 (0.543)
ch_sp400_evql5	−0.743 (0.838)	ch_spc_evql5	2.653*** (0.491)	ch_evql2_evql5	0.526 (1.004)	ch_evql3_evql5	−0.392 (0.413)
ch_evql4_evql5	−0.044 (0.560)	ch_evql5	0.154 (0.167)	ch_sq_tax	−0.047 (0.125)	ch_sp100_tax	0.645*** (0.071)
ch_sp200_tax	−0.047 (0.094)	ch_sp300_tax	0.517*** (0.113)	ch_sp400_tax	0.486*** (0.136)	ch_spc_tax	0.472*** (0.046)
ch_evql2_tax	−0.753*** (0.126)	ch_evql3_tax	−0.097* (0.051)	ch_evql4_tax	−0.231*** (0.051)	ch_evql5_tax	−0.196*** (0.048)
ch_tax	0.276*** (0.037)						

Notes: Log-likelihood at convergence: −5372.510; adjusted ρ^2: 0.343. Cluster-robust standard errors are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively (two-tailed tests).

Table B.2
Ordinary least squares regression results based on means of conditions.

Parameter	Estimate
(Intercept)	3.258*** (0.360)
male	−0.495** (0.216)
age	0.018*** (0.007)

Notes: Residual standard error = 4.077 (df = 1433); R^2 = 0.008; Adjusted R^2 = 0.007; F-statistic = 5.892 (df = 2, 1433, p = 0.003). Standard errors are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively (two-tailed tests).

Table B.2 reports OLS results where the conditional mean estimates of marginal WTP for the "species remains" attribute, derived from the random parameters logit model, are regressed on gender and age.

Data availability

The main function is presented in the appendix.

References

Börger, T., Ngoc, Q.T.K., Kuhfuss, L., Hien, T.T., Hanley, N., Campbell, D., 2021. Preferences for coastal and marine conservation in vietnam: Accounting for differences in individual choice set formation. Ecol. Econom. 180, 106885.

Campbell, D., 2007. Willingness to pay for rural landscape improvements: Combining mixed logit and random-effects models. J. Agric. Econ. 58 (3), 467–483.

Campbell, D., Hutchinson, W.G., Scarpa, R., 2009. Using choice experiments to explore the spatial distribution of willingness to pay for rural landscape improvements. Environ. Plan. A 41 (1), 97–111.

Campbell, D., Scarpa, R., Hutchinson, W.G., 2008. Assessing the spatial dependence of welfare estimates obtained from discrete choice experiments. Lett. Spat. Resour. Sci. 1 (2–3), 117–126.

Chappell, H.W., Greene, W., Harris, M.N., Spencer, C., 2022. Uncertainty and the Bank of England's MPC. J. Money Credit. Bank. 54 (4), 825–858.

Czajkowski, M., Budziński, W., Campbell, D., Giergiczny, M., Hanley, N., 2016. Spatial heterogeneity of willingness to pay for forest management. Environ. Resour. Econ. 68 (3), 705–727.

Daly, A.J., Hess, S., Train, K.E., 2012. Assuring finite moments for willingness to pay in random coefficient models. Transportation 39 (1), 19–31.

Dekker, T., Bansal, P., Huo, J., 2025. Revisiting mcfadden's correction factor for sampling of alternatives in multinomial logit and mixed multinomial logit models. Transp. Res. Part B: Methodol. 192, 103129.

Greene, W.H., Hensher, D.A., Rose, J.M., 2005. Using classical simulation-based estimators to estimate individual WTP values: A mixed logit case study of commuters. In: Scarpa, R., Alberini, A. (Eds.), Applications of Simulation Methods in Environmental and Resource Economics. Springer, pp. 17–33.

Heckman, J.J., 1979. Sample selection bias as a specification error. Econ.: J. Econ. Soc. 153–161.

Hess, S., 2010. Conditional parameter estimates from mixed logit models: Distributional assumptions and a free software tool. J. Choice Model. 3 (2), 134–152.

Hess, S., Palma, D., 2025. Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. Package version 0.3.5.

Hess, S., Train, K.E., Polak, J.W., 2006. On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a mixed logit model for vehicle choice. Transp. Res. Part B: Methodol. 40 (2), 147–163.

Johnston, R.J., Ramachandran, M., 2014. Modeling spatial patchiness and hot spots in stated preference willingness to pay. Environ. Resour. Econ. 59 (3), 363–387.
Krinsky, I., Robb, A.L., 1986. On approximating the statistical properties of elasticities. Rev. Econ. Stat. 715–719.
McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. J. Appl. Econometrics 15 (5), 447–470.
Murphy, K.M., Topel, R.H., 1985. Estimation and inference in two-step econometric models. J. Bus. Econom. Statist. 3 (4), 370–379.
Newey, W.K., 1984. A method of moments interpretation of sequential estimators. Econom. Lett. 14 (2–3), 201–206.
Pagan, A., 1984. Econometric issues in the analysis of regressions with generated regressors. Internat. Econom. Rev. 221–247.
Revelt, D., Train, K., 2000. Customer-specific taste parameters and mixed logit: Households' choice of electricity supplier. Working Paper E00-274. University of California, Berkeley, Department of Economics, Berkeley, CA, USA.
Richter, L., Pollitt, M.G., 2018. Which smart electricity service contracts will consumers accept? The demand for compensation in a platform market. Energy Econ. 72, 436–450.
Sandorf, E.D., Aanesen, M., Falk-Andersson, J., Furuseth, I.S., Hanley, N., Kaiser, B.A., Kourantidou, M., Navrud, S., Vondolia, G.K., Xuan, B.B., 2025. Exploring information and embedding effects on willingness-to-pay to control the invasive red king crab in Norway. Environ. Resour. Econ. 88 (9), 2529–2556.
Sandorf, E.D., Campbell, D., Hanley, N., 2017. Disentangling the influence of knowledge on attribute non-attendance. J. Choice Model. 24, 36–50.
Sarrias, M., 2020. Individual-specific posterior distributions from mixed logit models: Properties, limitations and diagnostic checks. J. Choice Model. 36, 100224.
Sarrias, M., Daziano, R., 2018. Individual-specific point and interval conditional estimates of latent class logit parameters. J. Choice Model. 27, 50–61.
Sillano, M., de Dios Ortúzar, J., 2005. Willingness-to-pay estimation with mixed logit models: Some new evidence. Environ. Plan. A 37 (3), 525–550.
Train, K.E., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, New York.
Vij, A., Hess, S., 2025. Posterior inference of attitude–behaviour relationships using latent class choice models. Available at SSRN: https://ssrn.com/abstract=5402434.