Random Utility Models with Skewed Random Components: The Smallest versus Largest Extreme Value Distribution

Richard T. Carson*, Derrick H. Sun†, and Yixiao Sun‡

September 4, 2026

Abstract

Most random utility models (RUMs) assume a right-skewed Largest Extreme Value Type I (LEVI) distribution for unobserved utility. We instead use its mirror image, the left-skewed Smallest Extreme Value Type I (SEVI) distribution. While equivalent in binary choice, these specifications progressively diverge as the choice set expands, resulting in substantially different estimates of preference parameters, average partial effects, market shares, and welfare measures.

Ranked by systematic utility, the alternatives split at any cutoff into a lagging group and a leading group. Because the SEVI upper tail is thin, an idiosyncratic shock rarely overturns a systematic lead: at every cutoff, the leading group takes more probability and the lagging group less, so the SEVI choice distribution first-order stochastically dominates the LEVI one. The same force shapes welfare. With three or more alternatives, the option value of a choice set is strictly higher under LEVI, and the gap is widest when the alternatives are evenly matched. Removing a systematic laggard costs less under SEVI, while, with four or more alternatives, removing a sufficiently dominant one costs more.

Conceptually, the models differ in whether idiosyncratic information is more likely to provide "hidden gems" (LEVI) or "hidden flaws" (SEVI). Across six samples—three from utility-maximization applications and three from cost-minimization applications—drawn from four standard empirical datasets, models with SEVI errors attain higher likelihood values than models with LEVI errors, making SEVI a serious candidate for the random components in discrete choice models.

Keywords: Conditional Logit, Gumbel Distribution, Independence of Irrelevant Alternatives, Largest Extreme Value Type I Distribution, Multinomial Logit, Reverse Gumbel Distribution, Skewness, Smallest Extreme Value Type I Distribution.

JEL codes: C10, C18, C25

*Department of Economics, UC San Diego. Email: rcarson@ucsd.edu

†Department of Mathematics and College of Computing, Data Science, and Society, UC Berkeley. Email: derrick.han.sun@gmail.com

‡Department of Economics, UC San Diego. Email: yisun@ucsd.edu

1 Introduction

Random utility models (RUMs) are a cornerstone of applied microeconomics and other fields such as marketing, health policy, and transportation research. Since the seminal work of McFadden (1973), the vast majority of RUMs have been formulated as conditional or multinomial logit models, or their generalizations (Hensher et al., 2015; Train, 2009). These models posit that the total utility of an alternative comprises two components: a systematic component parametrized by observable attributes, and an idiosyncratic component known only to the agent (Manski, 1977).

Models in the logit family assume that the random component follows the Largest Extreme Value Type I (LEVI) distribution, also known as the Gumbel distribution (Gumbel, 1958), which is continuous, unimodal, asymmetric, and right-skewed with a long *right* tail. Its exact mirror image is the Smallest Extreme Value Type I (SEVI) distribution, also known as the Reverse Gumbel distribution, which is likewise asymmetric, but left-skewed with a long *left* tail. Despite this theoretical symmetry, SEVI-based models have remained largely absent from the applied researcher's toolkit. This paper fills that gap by providing a general econometric framework for RUMs based on the SEVI distribution, studying the properties of this class of RUMs, and developing tools for estimation and inference.

The reasons for this near-complete absence from applied work are worth setting out. A small, scattered literature (Anderson and de Palma, 1999; Misra, 2005; Hu, 2005) has proposed SEVI-like reverse error structures, but it has left the impression that the model is a niche tool for minimization problems, that it changes little empirically relative to LEVI when fitted, and that it becomes computationally impractical once the choice set grows (we return to this literature in detail later in this section).

Each part of that impression turns out to be overstated. SEVI is best understood as a general-purpose distributional alternative to LEVI, not one confined to minimization problems. Its divergence from LEVI grows as the choice set expands, and its computational burden is manageable using Gosper's hack (Knuth, 2011, §7.1.3, Exercise 20), a technique borrowed from the computer science literature.

To demonstrate the empirical stakes of the LEVI versus SEVI modeling decision, consider a simple example. An agent (individual i) chooses among five pairs of running shoes ($j \in \{1, 2, 3, 4, 5\}$). Before trying them on, the individual rates each pair based on observable features such as brand and price. These ratings capture the systematic components, $\{V_j\}$, of a random utility model. For illustrative purposes, assume that the V_j's are known to be $\{0.25, 0.50, 0.75, 1.50, 2.00\}$.[1] Next, the agent tries on each pair of shoes, making an idiosyncratic judgment on comfort and fit. This judgment supplies the random component, ε_{ij}, which combines with V_j to yield total utility $U_{ij} = V_j + \varepsilon_{ij}$ for each pair. The pair chosen for purchase is the one with the highest U_{ij}.

Now suppose that n customers face this same choice, with identical and known systematic components $\{V_j\}_{j=1}^5$, but different random components ε_{ij}. Customers try on the shoes, thereby observing their own random components, and then decide which of the five pairs to purchase. The question can now be framed solely in terms of the choice probabilities (or, equivalently, market shares) for each of the five pairs of shoes. What are the predicted probabilities of choosing each of the $J = 5$ pairs when the random components, ε_{ij}, are drawn from (a) the standard LEVI

[1] The vector of systematic utilities can vary across individuals. Here, for simplicity, we assume that the V_j's are the same across all individuals.

distribution, (b) the standard SEVI distribution, or (c) $\mathcal{N}(0, \pi^2/6)$? We refer to the last as NORM, a normal distribution normalized to have the same variance as the standard LEVI and SEVI distributions.

The answer to this question is displayed in Figure 1. The results are striking, particularly for the alternatives with the smallest and largest systematic utilities. With a LEVI error, the alternative with the smallest systematic utility has a market share that is more than 130% higher than its SEVI counterpart and more than 20% higher than its NORM counterpart. This pattern continues for the alternatives with the second-smallest and third-smallest systematic utilities, with LEVI giving rise to market shares that are 75% and 38% higher than their SEVI counterparts, respectively. However, this pattern reverses around a 20% market share (i.e., $1/J$, the equal share across alternatives), above which SEVI yields the higher market share. For the alternative with the second-largest systematic utility, the SEVI market share is 11% higher than its LEVI counterpart and 6% higher than its NORM counterpart. The pattern of a higher market share under SEVI intensifies for the alternative with the largest systematic utility. Here, SEVI leads to a market share 21% higher than its LEVI counterpart and 16% higher than its NORM counterpart.

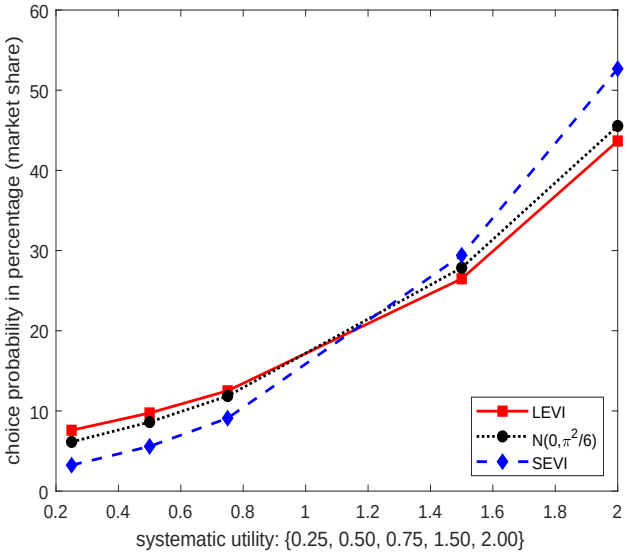


Figure 1: Market shares for all five alternatives in an RUM, with the systematic utilities held the same across the three distributions of the random utility components

Differences of this magnitude matter for common empirical uses. For instance, if alternative 1 were a new product, the use of a LEVI-based prediction when the true random component is SEVI would more than double the predicted demand (7.6% vs. 3.2% market shares). To the extent that new-product forecasts are systematically optimistic, SEVI random components offer one possible explanation. On the other hand, a LEVI-based prediction for the market leader, when the true random component is SEVI, could substantially underestimate its market share (43.7% vs. 52.7%). Either of these prediction errors could have important implications for production, profit, and antitrust analysis.

The driving force behind these differences is the skewness of the random component, a fundamental property that has been largely overlooked in the multinomial choice literature. From

the econometrician's viewpoint, the central question in our simple example is whether the unobserved utility for a randomly chosen pair of shoes is more prone to large positive surprises or to large disappointments. Because SEVI is the exact mirror image of LEVI, the two distributions share the same variance and the same excess kurtosis. Once recentered to a common mean, the only difference in their first four moments is the sign of their skewness, so the comparison drawn throughout this paper isolates the effect of skewness alone.

If the random component ε_{ij} exhibits a heavier upper tail than its lower tail, that is, if the distribution is prone to occasionally producing exceptionally high outlier values, then the random component is right-skewed. The classic exemplar in this scenario is the LEVI distribution, and the canonical model is the conditional logit. If the upper and lower tails are symmetric, making large and small outlier values equally likely, the normal distribution is the exemplar, and the multinomial probit is the canonical model. Conversely, if the random component exhibits a heavier lower tail than its upper tail, that is, if the distribution is prone to occasionally producing exceptionally small outlier values rather than large ones, then the random component is left-skewed. A well-established canonical model for this left-skewed case remains absent. We propose employing the SEVI distribution and its corresponding model as the natural counterpart to the standard LEVI formulation.

Fitting an RUM based on LEVI random components to data generated with SEVI random components, and vice versa, results in substantially biased estimates of the preference parameters. Even when the systematic components carry identical preference parameters, quantities routinely relied upon in decision making, such as elasticities and average partial effects, differ materially between the two models. Standard generalizations of the multinomial and conditional logit models, such as those allowing for latent classes or random parameters, do not address issues that arise when a LEVI random component is incorrectly assumed. SEVI-based analogues of those generalizations can be constructed to handle the same departures, such as correlation across alternatives and preference heterogeneity.

While we make no claim that the SEVI assumption is always or even generally preferable to that of LEVI, the SEVI model has clear empirical relevance. When examining six samples drawn from four datasets in standard econometric textbooks and well-known papers, we find that the model with SEVI errors attains a higher likelihood than the model with LEVI errors in every one of them, and the highest likelihood among the three specifications in all but one. Three of these samples are utility-maximization applications and three are cost-minimization applications; the model with SEVI errors attains the higher likelihood in both groups. For any particular application, the skewness of the random component is ultimately an empirical matter.

Most of the paper studies the theoretical properties of the SEVI model and its estimation, inference, and prediction. We preview several of these results.

First, in a SEVI-based RUM, the observed choice and the choice that maximizes the systematic utility are more likely to coincide than in the LEVI case, because SEVI random components have less influence over which alternative is selected. Conversely, they play a larger role than their LEVI counterparts in determining the worst option.[2] Because alternatives under SEVI are less likely to be "saved" by a large random shock, the model assigns more mass to the systematic leader and less to the laggard, properties we label "Leader-Takes-More" and "Laggard-Takes-Less." This ordering in fact holds at every cutoff in systematic utility, not only at the two extreme ones. Split

[2]One way to conceptualize the role of the random components, consistent with this observation, is that an agent with SEVI random components seeks idiosyncratic reasons to reject alternatives ("hidden flaws"), whereas under LEVI the agent aims to find a high-quality idiosyncratic match component to accept an alternative ("hidden gems").

the alternatives at any such cutoff into a lagging group and a leading group. The lagging group then takes less under SEVI and the leading group takes more, so that the SEVI choice distribution first-order stochastically dominates the LEVI one.

Second, the two models have different implications for option values and welfare analysis. With two alternatives, they deliver exactly the same option value, whatever the systematic utilities. With three or more, the option value is strictly higher under LEVI, whose heavier upper tail makes the right to choose after the shocks are observed more valuable. The same force governs the welfare loss from removing an alternative: the loss from removing a systematic laggard is lower under SEVI, since minor alternatives carry less option value than under LEVI, while, once there are four or more alternatives, removing a sufficiently dominant one costs more under SEVI.

Third, misspecifying the skewness does not damage every object of interest equally. When a likelihood based on the wrong error distribution is fitted, the estimated coefficients converge to the truth up to a common scale factor under a linearity condition on the covariates, so ratios of coefficients, and hence willingness-to-pay measures, remain consistent even though levels, elasticities, and average partial effects do not. Because the two models carry the same number of parameters, choosing between them by likelihood requires no tuning: the likelihood ranking coincides with ranking by AIC or BIC.

Fourth, the choice probabilities, the surplus function, and the compensating variation in a SEVI model all have intuitive, closed-form representations. In addition, a SEVI model is identified under the same conditions that identify the corresponding LEVI model, and no new identification difficulty arises. Estimation is standard as well: consistency of the MLE requires only a finite first moment of the covariates, the very condition that logit already imposes. Nor is the model much more costly to fit. With the choice subsets enumerated efficiently, the SEVI likelihood costs about as much as the LEVI likelihood in our simulations with up to roughly a dozen alternatives, and no more than four times as much in our applications, far less in both cases than the multinomial probit.

Finally, the structure of the choice probability in the SEVI model is considerably richer than in the LEVI model, where the choice probability depends on other alternatives solely through the sum of the exponentiated systematic components over all alternatives. In the SEVI model, the choice probability for a given alternative can be seen as an aggregation involving all choice subsets in which the alternative occurs. This subset structure, distinctive to SEVI, is also why the model does not impose the independence of irrelevant alternatives (IIA) property: it produces flexible substitution patterns even though the random components remain iid across the alternatives.

Historically, the "reverse" error structure has surfaced only sporadically and has never taken hold in applied work. The SEVI model was first proposed by Anderson and de Palma (1999) . Their paper lays out some of the model's key theoretical properties, such as the lack of the IIA property and the closed forms for both the choice probabilities and the surplus function, which enter below as Lemmas 1 and 2. It offers no empirical example, framing the model instead in an industrial organization context involving cost-minimizing behavior. Misra (2005) uses the model in an empirical marketing context, framing it as a regret-minimization model.[3] In the same year, Hu (2005) independently rediscovers the model, referring to it as "Minlogit" and providing another empirical application in a marketing context. Hu treats the model as simply having a different random component specification and states that it lacks a closed-form solution, without

[3]We are grateful to Timothy Brathwaite (`https://timothyb0912.github.io/blog/about/`) for bringing both Anderson and de Palma (1999) and Misra (2005) to our attention and for his helpful comments on our paper. At the time this paper was originally written, we were not aware of the sparse literature on reverse discrete choice models arising from cost or disutility minimization problems.

citing Anderson and de Palma (1999), who had already derived one. Like Misra (2005), Hu finds little difference between the LEVI- and SEVI-based estimates in an empirical application involving three choice alternatives, and reports that estimation becomes impractical once the number of alternatives is appreciable. Taken together, this literature has left an incomplete impression of the model's scope and practicality.[4]

Our paper provides the first systematic comparison of LEVI and SEVI within a common utility-maximization framework, and one that rests on theorems rather than on fitted examples. The orderings established below hold at every configuration of systematic utilities, so the differences between the two models are properties of the two error distributions themselves and not artifacts of a particular dataset or parameter estimate. Read in this light, the earlier finding of "little difference" is not evidence against the two specifications being distinct. It comes from applications with three alternatives, and three is close to the least favorable case for detecting a difference: the two models coincide exactly at $J = 2$, and at $J = 3$ the margin that a systematic leader must overcome is, in expectation, still the same under both (Section S.G.2). Separation sets in beyond $J = 3$: several of the orderings below turn strict only at $J \geq 3$ or $J \geq 4$, and the gap widens thereafter in our simulations.

The rest of the paper is organized as follows. Section 2 studies the structure of the choice probabilities in a SEVI model. Section 3 presents a closed-form expression for the surplus function and compares it with its LEVI counterpart, then applies it to the compensating variation for a price change and for the removal of an alternative, and to the identification of the SEVI model. Section 4 covers estimation and model selection, while Section 5 presents simulation evidence. Section 6 applies the model to data from textbooks and well-known papers, comparing goodness of fit across LEVI, SEVI, and NORM. The final section concludes and discusses directions for future research. The main appendix proves Theorems 1 and 2 and Corollary 1. The remaining proofs are in the supplementary appendix, along with additional materials and figures.

2 The Model and the Choice Probabilities

2.1 The SEVI-based RUM

We consider a probabilistic choice model based on utility maximization. We assume that the utility of individual i from alternative j is given by

$$U_{ij}(\beta_0) = V_{ij}(\beta_0) + \varepsilon_{ij}, \;\; j = 1, 2, \ldots, J, \tag{1}$$

where J is the number of alternatives, $V_{ij}(\beta_0)$ is the systematic utility component observable to the econometrician, ε_{ij} captures the unobservable random utility component, and $\{\varepsilon_{ij}\}$ is independent of $\{V_{ij}(\beta_0)\}$. Individuals choose the alternative that maximizes their total utility. Let Y_i denote the choice of individual i. Then

$$Y_i = \arg \max_{j=1,\ldots,J} U_{ij}(\beta_0).$$

For the random utility component, we assume that ε_{ij} is iid over $j = 1, 2, \ldots, J$ and follows the *Smallest* Extreme Value Type I (SEVI) distribution. The CDF of the standard SEVI distribution

[4]Uptake of this literature has been limited, and few of the papers citing it are empirical applications. More recently, there has been some interest in using the SEVI distributional assumption for a different type of minimization problem, namely picking the worst alternative out of a set, as is commonly done in a best-worst scaling choice task (Collins and Rose, 2013; Delle Site et al., 2019).

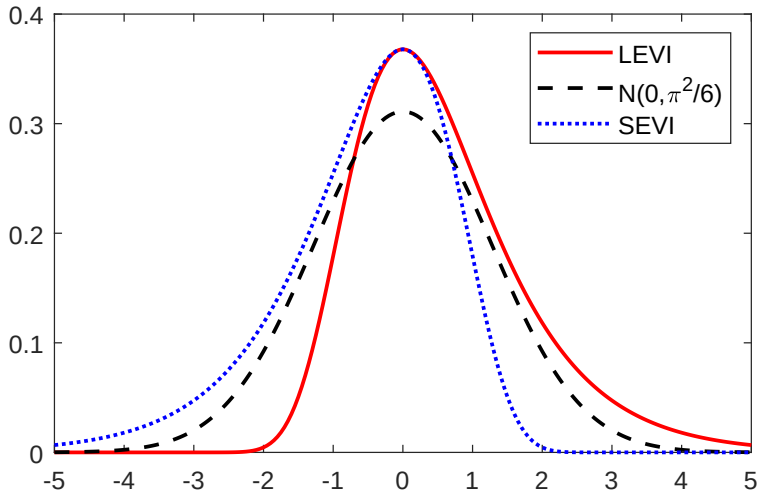


Figure 2: Probability density functions of the LEVI, NORM, and SEVI distributions

is $F(a) = 1 - \exp(-\exp(a))$, with the corresponding pdf:

$$f(a) = F'(a) = \exp(a - \exp(a)). \tag{2}$$

This distributional assumption is in contrast to that in standard logit models, where the latent utility is given by $\tilde{U}_{ij}(\beta_0) = V_{ij}(\beta_0) + \tilde{\varepsilon}_{ij}$, and the random utility component $\tilde{\varepsilon}_{ij}$ follows the standard Largest Extreme Value Type I (LEVI) distribution with CDF $\tilde{F}(a) = \exp(-\exp(-a))$. These two forms of extreme-value distributions can be regarded as mirror images, since $(\varepsilon_{i1}, \ldots, \varepsilon_{iJ})'$ and $-(\tilde{\varepsilon}_{i1}, \ldots, \tilde{\varepsilon}_{iJ})'$ share the same distribution. For easy reference, we call a random component following the SEVI distribution a SEVI error and the resulting RUM the SEVI model. The same nomenclature applies to the LEVI case, and a LEVI model is the standard logit model.[5]

Figure 2 illustrates the pdfs of the LEVI and SEVI distributions alongside a normal benchmark. While LEVI exhibits strong right-skewness, SEVI exhibits the opposite behavior. Under LEVI, the probability that a random draw exceeds x is $1 - \exp(-\exp(-x))$, which is of order $\exp(-x)$ and decays exponentially as $x \to +\infty$. In contrast, the corresponding probability under SEVI is $\exp(-\exp(x))$, which decays *double-exponentially* as $x \to +\infty$. Consequently, the chance of a very large upward "surprise" is substantially smaller under SEVI. This makes the SEVI distribution appropriate for modeling choice environments characterized by "downside risk," where alternatives are more likely to suffer from rare but severe "hidden flaws" than to benefit from the rare "hidden gems" implied by standard models.

This notion of random-component skewness is inherently conditional on the alternative attributes and agent characteristics observed by the econometrician. In a linear latent utility model, for instance, ε_{ij} contains unobserved shocks such as $(\tilde{X}_{ij} - E\tilde{X}_{ij})\tilde{\beta}_{0,x}$ for a scalar attribute $\tilde{X}_{ij}$ known to the individual but not to the econometrician. If $\tilde{X}_{ij}$ has negative skewness and $\tilde{\beta}_{0,x} > 0$, then $(\tilde{X}_{ij} - E\tilde{X}_{ij})\tilde{\beta}_{0,x}$ will also have negative skewness, so ε_{ij} may be left-skewed, for which the SEVI distribution is a natural candidate. Consequently, the inclusion or exclusion of a particular covariate may change the skewness of the random component. More generally, specification errors in the systematic utility component effectively become part of the random component. Since

[5]Throughout the paper, SEVI means the standard SEVI distribution, with pdf given in (2) and variance $\pi^2/6$. A similar comment applies to the LEVI case.

such errors can manifest as left-skewed, symmetric, or right-skewed, the SEVI distribution is at least as plausible a candidate for the random component as the LEVI distribution.

Our choice between LEVI and SEVI is driven by the skewness of the unobserved factors, not by any extreme-value micro-foundation. While the standard LEVI assumption is sometimes justified via the Fisher–Tippett–Gnedenko theorem (see, for example, de Haan and Ferreira (2007)) by assuming that ε_{ij} represents the maximum of many unobserved sub-attributes, we do not require ε_{ij} to be the minimum of sub-attributes to justify using SEVI. In our framework, the consumer's decision rule is to maximize total utility; this top-level maximization is an economic behavioral assumption that does not dictate the distributional shape of the primitive alternative-specific shocks.

Furthermore, the SEVI distribution sits outside the generalized extreme value class of choice models (Train, 2009, Ch. 4), every member of which has Gumbel marginals. While one could in principle employ the negative Weibull, a special case of the generalized extreme value *distribution*, to capture negative skewness, doing so inherently truncates the support. By utilizing the exact reflection of the standard LEVI, our framework captures left-skewed unobservables while maintaining support on the entire real line, thereby avoiding the arbitrary hard bounds and additional shape parameters that would otherwise complicate the econometric analysis.

Finally, while the LEVI and SEVI distributions have different means, γ (the Euler–Mascheroni constant) and $-\gamma$, respectively, this difference has no effect on choices, as the arg max is invariant to any constant added to all alternatives. To make the models directly comparable, we can recenter the error terms by specifying $U_{ij}(\beta_0) = V_{ij}(\beta_0) + \tilde{e}_{ij}$ for $\tilde{e}_{ij} = e_{ij} - E(e_{ij})$, where e_{ij} follows either the LEVI or SEVI distribution. Now the random components in both models have mean zero, so the distinction between LEVI and SEVI is one of shape, not of the mean. In fact, since the recentered LEVI and SEVI errors also share the same variance ($\pi^2/6$) and excess kurtosis (12/5), the only difference in their first four moments is the sign of their skewness.

2.2 Choice Probabilities in the SEVI-based RUM

In a SEVI-based RUM, as in the LEVI-based model, choice probabilities have closed-form expressions, albeit with a more complex combinatorial structure. Let $V_i := (V_{i1}(\beta_0), \ldots, V_{iJ}(\beta_0))'$ denote the vector of systematic utilities. To simplify notation, we suppress dependence on β_0. We then write V_{ij} and V_i for $V_{ij}(\beta_0)$ and $V_i(\beta_0)$, respectively. We adopt this same convention for realized values v_{ij} and v_i, and for other quantities such as U_{ij} and U_i.

Following Yellott (1977), we also define the scale values:

$$S_{ij}(\beta_0) = \exp(V_{ij}(\beta_0)), \; s_{ij}(\beta_0) = \exp(v_{ij}(\beta_0)), \; S_i = (S_{ij})_{j=1}^J, \; s_i = (s_{ij})_{j=1}^J.$$

Lemma 1 *Assume that $U_{ij} = V_{ij} + \varepsilon_{ij}$, $j = 1, 2, \ldots, J$, where $\varepsilon_i := (\varepsilon_{i1}, \ldots, \varepsilon_{iJ})'$ is independent of V_i, and the elements of ε_i are independent and identically distributed, each following the SEVI distribution. Then, under utility maximization, the probability of individual i choosing alternative j is*

$$P^{\text{SEVI}}(Y_i = j|v_i) := P^{\text{SEVI}}(Y_i = j|V_i = v_i) = \int_0^1 \prod_{k \in \mathcal{J}_{-j}} \left[1 - u^{s^i_{j|k}}\right] du \tag{3}$$

$$= 1 + \sum_{\ell=1}^{J-1} (-1)^\ell \sum_{|\mathcal{S}|=\ell, \mathcal{S} \subseteq \mathcal{J}_{-j}} \frac{1}{1 + \sum_{k \in \mathcal{S}} s^i_{j|k}}, \tag{4}$$

where $s^i_{j|k} = s_{ij}/s_{ik}$, the inner sum runs over the subsets $\mathcal{S}$ of $\mathcal{J}_{-j}$ with ℓ elements, and for $\mathcal{J} = \{1, \ldots, J\}$, $\mathcal{J}_{-j} = \mathcal{J} \setminus \{j\}$ is the subset of alternatives excluding the j-th alternative.

Lemma 1 provides two representations of the choice probability under SEVI. It is the integral (3), not the alternating sum (4), that carries the analysis: the proofs of Theorems 1 and 2, of Corollary 1, and of the envelope bound underlying Theorem 4 all operate on that representation.[6]

For comparison, the LEVI choice probability is

$$P^{\text{LEVI}}(Y_i = j|v_i) := P^{\text{LEVI}}(Y_i = j|V_i = v_i) = \frac{1}{1 + \sum_{k \in \mathcal{J}_{-j}} \exp(v_{ik} - v_{ij})} = \frac{1}{1 + \sum_{k \in \mathcal{J}_{-j}} s^i_{k|j}}, \tag{5}$$

where $s^i_{k|j} := s_{ik}/s_{ij} = 1/s^i_{j|k}$. Since $S_i = \exp(V_i)$ is one-to-one, we may condition on $S_i = s_i$ instead of $V_i = v_i$ in all probability calculations. In both cases, moreover, only the pairwise ratios of the scale values $\{s_{ij}\}_{j=1}^J$ appear, so the choice probabilities are unchanged when every scale value is multiplied by a common positive constant. However, the LEVI-based probability takes the familiar simple form, while the SEVI-based probability exhibits a richer combinatorial structure involving alternating sums over all nonempty subsets of competing alternatives.

For $J = 2$, the SEVI and LEVI models yield identical probabilities:

$$P^{\text{SEVI}}(Y_i = 1|v_i) = 1 - \frac{1}{1 + s^i_{1|2}} = \frac{s^i_{1|2}}{1 + s^i_{1|2}} = \frac{1}{1 + s^i_{2|1}} = P^{\text{LEVI}}(Y_i = 1|v_i).$$

This equivalence arises because the difference between two iid SEVI draws follows the same logistic distribution as the difference between two iid LEVI draws.

When $J > 2$, the choice probabilities diverge. Consider the case with $J = 3$. It follows from Lemma 1 that

$$P^{\text{SEVI}}(Y_i = 1|v_i) = 1 - \frac{1}{1 + s^i_{1|2}} - \frac{1}{1 + s^i_{1|3}} + \frac{1}{1 + s^i_{1|2} + s^i_{1|3}}. \tag{6}$$

This is different from the LEVI choice probability, which is equal to

$$P^{\text{LEVI}}(Y_i = 1|v_i) = \frac{1}{1 + s^i_{2|1} + s^i_{3|1}}.$$

To understand the formula for $P^{\text{SEVI}}(Y_i = 1|v_i)$, we can examine $1 - P^{\text{SEVI}}(Y_i = 1|v_i)$, which represents the probability that alternative 1 is *not* chosen. The latter event can be expressed as a union of two events $\{U_{i1} \leq U_{i2}\} \cup \{U_{i1} \leq U_{i3}\}$, indicating that alternative 1 is dominated by either alternative 2 or alternative 3. The probability of this union is

$$\begin{aligned}
&\Pr[\{U_{i1} \leq U_{i2}\} \cup \{U_{i1} \leq U_{i3}\}] \\
&= \Pr[U_{i1} \leq U_{i2}] + \Pr[U_{i1} \leq U_{i3}] - \Pr[\{U_{i1} \leq U_{i2}\} \cap \{U_{i1} \leq U_{i3}\}] \\
&= \Pr[-U_{i1} \geq -U_{i2}] + \Pr[-U_{i1} \geq -U_{i3}] - \Pr[\{-U_{i1} \geq -U_{i2}\} \cap \{-U_{i1} \geq -U_{i3}\}].
\end{aligned}$$

[6]The analytical structure in (4) parallels what Anderson and de Palma (1999) established for cost-minimizing choice with LEVI errors, a problem equivalent to utility-maximizing choice with SEVI errors; there too the choice probabilities are built from logit-type terms (see also Misra (2005)). We record the result in the notation and in the integral form used throughout this paper, and provide the derivation in the supplementary appendix. The novelty lies not in the representations but in their consequences, which the rest of the paper develops.

Each of the three probabilities above corresponds to a choice probability in a conventional LEVI model with flipped systematic utilities. Substituting the familiar LEVI formulae yields equation (6), and Lemma 1 generalizes this argument.

When $v_{ij} = v_{ik}$ for all j and k, the alternatives are indistinguishable, and the choice probability simplifies to

$$P^{\text{SEVI}}(Y_i = j|v_i) = \int_0^1 (1-u)^{J-1}\, du = \frac{1}{J}.$$

In this case, the SEVI model yields the same result as the LEVI model. On the other hand, when v_{ij} increases and dominates all other v_{ik} for $k \neq j$, $P^{\text{SEVI}}(Y_i = j|v_i)$ approaches 1. The choice probability is thus a type of softmax function mapping $\mathbb{R}^J$ into the interior of the unit simplex. For example, it maps $(1, 2, 8)$ into $\left(4.24 \times 10^{-4}, 2.29 \times 10^{-3}, 0.9973\right)$, assigning nearly all probability mass to the maximal element. While this behavior is qualitatively similar to that in the LEVI model, the SEVI softmax is "sharper," as shown by the following two theorems, the second of which states the comparison as a stochastic ordering over the whole choice distribution.

Theorem 1 *Let $v_1, v_2, \ldots, v_J$ be the systematic utilities.*

(a) If $v_1 > \max(v_2, \ldots, v_J)$ and $J > 2$, then $P^{\text{SEVI}}(Y = 1|v) > P^{\text{LEVI}}(Y = 1|v)$.

(b) If $v_1 < \min(v_2, \ldots, v_J)$ and $J > 2$, then $P^{\text{SEVI}}(Y = 1|v) < P^{\text{LEVI}}(Y = 1|v)$.

(c) Fix $v_2, \ldots, v_J$ and let $v_1 \to -\infty$, so that $s_1 \to 0$. Then

$$P^{\text{LEVI}}(Y = 1|v) = \left(\sum_{k=2}^{J} s_k\right)^{-1} s_1\left(1 + o\left(1\right)\right),$$

$$P^{\text{SEVI}}(Y = 1|v) = (J-1)!\left(\prod_{k=2}^{J} s_k^{-1}\right) s_1^{J-1}\left(1 + o\left(1\right)\right). \tag{7}$$

Theorem 1 shows that the SEVI model is more sensitive to systematic utility differences than the standard logit. In Theorem 1(a), the leader "takes more" because the double-exponentially decaying upper tail of the SEVI distribution makes it rare for a stochastic shock to overcome a systematic lead. Conversely, in Theorem 1(b), the laggard "takes less" for the same reason: unobserved shocks are less likely to lift a laggard under SEVI than under LEVI. This two-sided effect indicates that the SEVI model stays closer than the LEVI model to the choice that maximizes the systematic utility. Figure 1 in the introduction illustrates both cases at its two ends: the leading alternative's market share is 21% higher under SEVI than under LEVI, and the lagging alternative's is less than half as large.

Theorem 1(c) quantifies Theorem 1(b) in the limit. Both choice probabilities vanish as alternative 1 becomes negligible, but at different rates: the SEVI probability is of order s_1^{J-1} and the LEVI probability of order s_1, so their ratio tends to zero for $J > 2$. A systematic laggard is thus discounted not merely more heavily under SEVI but at a polynomial rate that rises with the size of the choice set.

For $J = 3$, the difference in the choice probabilities can be evaluated in closed form. This leads to a sharper result than Theorem 1(a) and (b): $P^{\text{SEVI}}(Y = 1|v) > P^{\text{LEVI}}(Y = 1|v)$ if and only if v_1 exceeds the average of v_2 and v_3. One rival may thus be stronger than alternative 1, provided the other is correspondingly weaker. See Proposition S.2 in the supplementary appendix for details.

Theorem 1(a) and (b) compare the two models at a single alternative, the one ranked above or below all others. When the whole ranking of the systematic utilities is known, the comparison extends to every threshold at once.

Theorem 2 *Let $v_1, v_2, \ldots, v_J$ be the systematic utilities for $J > 2$ alternatives, labeled so that $v_1 \leq v_2 \leq \cdots \leq v_J$.*

(a) For every $j = 1, \ldots, J-1$,

$$P^{\text{SEVI}}(Y \leq j|v) \leq P^{\text{LEVI}}(Y \leq j|v).$$

(b) If $v_1 < v_J$, the inequality in part (a) is strict for every $j = 1, \ldots, J-1$.

Theorem 2 says that the choice distribution under SEVI first-order stochastically dominates that under LEVI when alternatives are ordered by systematic utility. Consider splitting the alternatives at the cutoff v_j into a lagging group, the j lowest-ranked ones, and a leading group, the rest. The lagging group takes less probability under SEVI than under LEVI, and the leading group correspondingly takes more. Because this holds for every $j = 1, \ldots, J-1$, the leaders take more and the laggards less group by group, not only at the two ends of the ranking. The comparison is between groups rather than between individual alternatives, and an alternative of intermediate rank may receive less probability under SEVI even though it is not the lowest-ranked. This is the formal content of the statement that SEVI choices align more closely with systematic utility maximization. Theorem 2(b) shows that strictness is all-or-nothing: unless the alternatives are all identical in systematic utility, the inequality is strict at every threshold v_j, so that a tie at the splitting index, $v_j = v_{j+1}$, does not by itself cost strictness.

Theorem 1(a) and (b) are the extreme cases. Under the assumption of either part, the top and bottom systematic utilities differ, so Theorem 2(b) applies and recovers a strict ordering. Taking $j = J-1$ in Theorem 2 and passing to complements returns Theorem 1(a), since alternative J is then the highest-ranked. Taking $j = 1$ in Theorem 2 returns Theorem 1(b), since alternative 1 is the lowest-ranked. "Leader-Takes-More" and "Laggard-Takes-Less" are thus the two ends of a single stochastic ordering, and the new content of Theorem 2 lies at the intermediate thresholds $1 < j < J-1$, a nonempty range only when $J \geq 4$.

Although Theorem 1(a) and (b) are recovered as special cases, Theorem 2 is not an independent proof of Theorem 1: the proof of Theorem 2 invokes Theorem 1(a) on auxiliary models and Theorem 1(b) on the model itself. There is no circularity, because both parts of Theorem 1 are proved on their own for every $J > 2$ without using Theorem 2.

Theorem 2 compares the two models threshold by threshold. Averaging any nondecreasing function of an alternative's rank under the choice probabilities collapses those $J-1$ comparisons into a single scalar inequality. For $\mathcal{M} \in \{\text{LEVI}, \text{SEVI}\}$, collect the choice probabilities in the vector

$$P^{\mathcal{M}}(v) := \left(P^{\mathcal{M}}(Y = 1|v), \ldots, P^{\mathcal{M}}(Y = J|v)\right)',$$

write $\langle x, y \rangle := \sum_{j=1}^{J} x_j y_j$ for the Euclidean inner product and $\|x\|^2 := \langle x, x \rangle$, and let

$$E^{\mathcal{M}}(v_Y|v) := \left\langle P^{\mathcal{M}}(v), v \right\rangle$$

denote the expected systematic utility of the alternative chosen, the average of v_Y over repeated decisions at the same v.

Corollary 1 *Let $v_1 \leq v_2 \leq \cdots \leq v_J$ with $J > 2$. Then:*

(a) $\left\langle P^{\text{SEVI}}(v), w \right\rangle \geq \left\langle P^{\text{LEVI}}(v), w \right\rangle$ for every $w \in \mathbb{R}^J$ with $w_1 \leq w_2 \leq \cdots \leq w_J$;

(b) $E^{\text{SEVI}}(v_Y|v) = \left\langle P^{\text{SEVI}}(v), v \right\rangle \geq \left\langle P^{\text{LEVI}}(v), v \right\rangle = E^{\text{LEVI}}(v_Y|v)$;

(c) $\left\|P^{\text{SEVI}}(v)\right\|^2 \geq \left\|P^{\text{LEVI}}(v)\right\|^2$.

If $v_1 < v_J$, then the inequalities in (b) and (c) are strict. The inequality in (a) is then also strict for every such w with $w_{j+1} > w_j$ for at least one j.

Part (b) is Part (a) at $w = v$; Part (c) follows by applying Part (a) at $w = P^{\mathcal{M}}(v)$ for each of the two models in turn, the latter being nondecreasing because an alternative with a higher systematic utility has a higher choice probability. Part (b) sharpens the sense in which SEVI choices track the systematic ranking: the alternative actually chosen carries a higher systematic utility on average under SEVI than under LEVI.

Written as an inner product, Part (b) also has a regression reading. Encode the decision as the vector of indicators $\iota_Y := (1\{Y = 1\}, \ldots, 1\{Y = J\})'$, whose expectation is $P^{\mathcal{M}}(v)$, and write $\bar{v} := J^{-1}\sum_{j=1}^{J} v_j$. Provided the v_j are not all equal, regressing the J coordinates of ι_Y on $v_1, \ldots, v_J$ with an intercept gives the slope

$$\hat{b} = \frac{\sum_{j=1}^{J}(v_j - \bar{v})\,1\{Y = j\}}{\sum_{j=1}^{J}(v_j - \bar{v})^2} = \frac{v_Y - \bar{v}}{\sum_{j=1}^{J}(v_j - \bar{v})^2},$$

so its mean is $E^{\mathcal{M}}(\hat{b}|v) = \left[E^{\mathcal{M}}(v_Y|v) - \bar{v}\right] / \sum_{j=1}^{J}(v_j - \bar{v})^2$, the same number one obtains by regressing the choice probabilities $P^{\mathcal{M}}(v)$ on v. Part (b) is exactly the statement that this mean slope is larger under SEVI: across alternatives, the choice indicator co-moves more strongly with the systematic utilities.

Part (c) says that the SEVI choice probabilities are more spread out around the equal share $1/J$: both vectors have coordinate mean $1/J$, so $\sum_{j=1}^{J}\left(P^{\mathcal{M}}(Y = j|v) - 1/J\right)^2 = \left\|P^{\mathcal{M}}(v)\right\|^2 - 1/J$, and Part (c) orders the left-hand side. Because $\left\|P^{\mathcal{M}}(v)\right\|^2$ is the Herfindahl index of the implied market shares, a SEVI analyst predicts a more concentrated market than a LEVI analyst at the same systematic utilities, strictly so whenever $v_1 < v_J$.

The results so far in this subsection leave one question open. Theorem 1(a) and (b) fix the comparison only at the alternative ranked above or below all others. Theorem 2 applies to every configuration once the alternatives are relabeled in order of systematic utility, but it compares the two models only at the cumulative probabilities, and Corollary 1 only at weighted combinations of them. The gap between the two models at a single alternative is a difference of two such cumulative gaps, so none of these results signs it at an intermediate rank, and an alternative of intermediate rank can indeed have a smaller choice probability under SEVI. Section S.H of the supplementary appendix simulates the two models under the data-generating processes (DGPs) of Section 5 and shows that the amplification found at the two extremes continues through the intermediate ranks: alternatives with a higher LEVI probability tend to gain under SEVI and those with a lower one tend to lose. This is a tendency across realizations of a particular DGP rather than a general rule for intermediate alternatives.

The mechanism behind all these comparisons can be located in the random components alone, before any systematic utility is specified. Let $M = \max_{j \neq 1}(\varepsilon_j - \varepsilon_1)$ be the largest advantage that any rival draws over alternative 1, the quantity a systematic lead must overcome for that alternative to be chosen; the choice of label is immaterial because the random components are iid. Section S.G.2 of the supplementary appendix shows that for $J \geq 4$ the mean of M is smaller under SEVI than under LEVI, increasingly so as J grows, so SEVI shocks are intrinsically less able to overturn the ranking set by the systematic utilities.

2.3 Patterns of Substitution and the Lack of IIA

In this subsection, we compute $\partial P^{\mathrm{SEVI}}(Y_i = j|v_i)/\partial v_{ik}$, which measures how a change in alternative k's systematic utility affects the probability of choosing alternative j. This derivative is

a building block for determining $\partial P^{\text{SEVI}}(Y_i = j|v_i)/\partial X_{ik}$, the marginal effect of changing an alternative's attributes on choice probabilities.

For notational economy, we let $P^{\text{SEVI}}_{ik} := P^{\text{SEVI}}(Y_i = k|v_i)$ and write $\mathcal{S}_\ell$ for a subset of the alternatives with ℓ elements. Each term of the alternating sum in Lemma 1 is itself a logit probability at reflected utilities, since $1/(1 + \sum_{m\in\mathcal{S}_\ell} s^i_{j|m}) = P^{\text{LEVI}}(Y_i = j|\mathcal{S}_\ell \cup \{j\}, -v_i)$, the probability of choosing alternative j from the subset $\mathcal{S}_\ell \cup \{j\}$ when the random components are iid LEVI and the systematic utilities are $-v_i$. Differentiating (4) term by term therefore gives

$$\frac{\partial P^{\text{SEVI}}_{ij}}{\partial v_{ij}} = -\sum_{\ell=1}^{J-1}(-1)^\ell \sum_{\mathcal{S}_\ell\subseteq\mathcal{J}_{-j}} P^{\text{LEVI}}(Y_i = j|\mathcal{S}_\ell \cup \{j\}, -v_i)\cdot P^{\text{LEVI}}(Y_i \neq j|\mathcal{S}_\ell \cup \{j\}, -v_i), \quad (8)$$

where $\sum_{\mathcal{S}_\ell\subseteq\mathcal{J}_{-j}}$ is short for the sum over the subsets of $\mathcal{J}_{-j}$ with ℓ elements, as in (4), and

$$\begin{aligned} P^{\text{LEVI}}(Y_i = j|\mathcal{S}_\ell \cup \{j\}, -v_i) &= \frac{\exp(-v_{ij})}{\sum_{m\in\mathcal{S}_\ell\cup\{j\}}\exp(-v_{im})}, \\ P^{\text{LEVI}}(Y_i \neq j|\mathcal{S}_\ell \cup \{j\}, -v_i) &= 1 - P^{\text{LEVI}}(Y_i = j|\mathcal{S}_\ell \cup \{j\}, -v_i). \end{aligned}$$

Furthermore, for $k \neq j$,

$$\frac{\partial P^{\text{SEVI}}_{ij}}{\partial v_{ik}} = \sum_{\ell=1}^{J-1}(-1)^\ell \sum_{\mathcal{S}_{\ell-1}\subseteq\mathcal{J}_{-(j,k)}} P^{\text{LEVI}}(Y_i = j|\mathcal{S}_{\ell-1}\cup\{j,k\}, -v_i)\cdot P^{\text{LEVI}}(Y_i = k|\mathcal{S}_{\ell-1}\cup\{j,k\}, -v_i), \quad (9)$$

where $\mathcal{J}_{-(j,k)}$ is the subset of $\mathcal{J}$ excluding j and k. In the above equation, the $\ell = 1$ term of the outer summation is understood to be

$$-\frac{\exp(-v_{ij})}{[\exp(-v_{ij}) + \exp(-v_{ik})]}\frac{\exp(-v_{ik})}{[\exp(-v_{ij}) + \exp(-v_{ik})]}.$$

Equation (9) reveals that

$$\frac{\partial P^{\text{SEVI}}_{ij}}{\partial v_{ik}} = \frac{\partial P^{\text{SEVI}}_{ik}}{\partial v_{ij}}$$

for any j and k. That is, the choice probabilities satisfy a symmetry property analogous to Slutsky symmetry in continuous choice models.

The cross-derivative $\partial P^{\text{SEVI}}_{ij}/\partial v_{ik}$ can be used to derive the score function, which is needed for computing the MLE and developing asymptotic theory. It can also be used to examine whether the IIA property holds for $J \geq 3$. For IIA to hold, changing one alternative's utility should not affect the ratio of probabilities between two others. Formally, the cross (semi)elasticity $\partial \log P^{\text{SEVI}}_{ij}/\partial v_{ik}$ should be the same for all $j \neq k$. That is,

$$\frac{\partial \log P^{\text{SEVI}}_{ij_1}}{\partial v_{ik}} = \frac{\partial \log P^{\text{SEVI}}_{ij_2}}{\partial v_{ik}}$$

for all $j_1 \neq k$ and $j_2 \neq k$. Equation (9) shows that this condition is not met. Alternatively, for IIA to hold, the probability ratio $P^{\text{SEVI}}_{ij_1}/P^{\text{SEVI}}_{ij_2}$ must depend only on the systematic utilities of alternatives j_1 and j_2. In the SEVI model, this ratio remains a function of the systematic utilities of all available alternatives (Anderson and de Palma, 1999).

In an additively separable RUM with $U_{ij} = V_{ij} + \varepsilon_{ij}$, as considered here, it is well known that the IIA property holds for any $J \geq 3$ if and only if the random component ε_{ij} follows the LEVI distribution up to location and scale (cf. Yellott (1977)). Since the SEVI distribution is not a location-scale transformation of the LEVI distribution, the failure of IIA in the SEVI model is an immediate consequence. This helps explain why substitution patterns in the SEVI model differ from those of the LEVI-based logit model.

To illustrate the violation of IIA in a SEVI model with $J = 3$, we consider adding a third alternative (alternative 3) to two initial alternatives (alternatives 1 and 2). Let s_1, s_2, and s_3 be the scale values of alternatives 1, 2, and 3, respectively. We fix $s_1 = 1$ ($v_1 = 0$) and let s_2 and s_3 vary. Define the difference in the probability ratios under the SEVI and LEVI models as

$$D_{\mathrm{PR}}(s_2, s_3) = \frac{P^{\mathrm{SEVI}}(Y = 2|v_1, v_2, v_3)}{P^{\mathrm{SEVI}}(Y = 1|v_1, v_2, v_3)} - \frac{P^{\mathrm{LEVI}}(Y = 2|v_1, v_2, v_3)}{P^{\mathrm{LEVI}}(Y = 1|v_1, v_2, v_3)}.$$

The subtracted term equals s_2/s_1, which is also the probability ratio that the two models share when only alternatives 1 and 2 are available, so D_{PR} measures how the SEVI ratio moves when alternative 3 is added. It thereby gauges the extent of IIA violation: if IIA held, D_{PR} would be identically zero. Some algebra shows that

$$D_{\mathrm{PR}}(s_2, s_3) = \frac{s_2 s_3^2 \left(s_2^2 - 1\right)}{\left(s_2 + s_3 + 2 s_2 s_3\right)\left(s_2 + s_3\right)} = \left(s_2^2 - 1\right)\left(\frac{2 s_2}{s_3} + \frac{s_2}{s_3^2} + \frac{1}{s_2} + \frac{2}{s_3} + 2\right)^{-1}.$$

The function $D_{\mathrm{PR}}(s_2, s_3)$ is not identically zero. When the third alternative is introduced, the market shares of alternatives 1 and 2 do not decrease proportionally in the SEVI model, indicating a violation of IIA.

An important consequence of the lack of IIA is the inconsistency of estimation based on a randomly sampled subset of alternatives. In LEVI-based logit models, parameters can be consistently estimated this way. See, for example, page 64 of Train (2009). In a SEVI model, this is not possible because

$$\frac{P^{\mathrm{SEVI}}(Y = 1|v_1, v_2, v_3)}{P^{\mathrm{SEVI}}(Y = 1|v_1, v_2, v_3) + P^{\mathrm{SEVI}}(Y = 2|v_1, v_2, v_3)} \neq P^{\mathrm{SEVI}}(Y = 1|v_1, v_2).$$

The lack of IIA implies that the relative probability of choosing between two alternatives depends on the full composition of the choice set, making "shortcut" sampling and estimation methods invalid for the SEVI specification.

3 The Surplus Function and Its Applications

A fundamental concept in discrete choice analysis is the surplus function, defined as:

$$W_J(v) = E\left[\max_{j \in \mathcal{J}} \{v_j + e_j\}\right],$$

where the expectation is taken with respect to the distribution of $\{e_1, \ldots, e_J\}$. Among the many uses of W_J, an important one rests on the fact that the choice probabilities are the gradient of the surplus function:

$$P(Y_i = j|V_i = v_i) = \frac{\partial W_J(v_{i1}, \ldots, v_{iJ})}{\partial v_{ij}}.$$

This is the Williams–Daly–Zachary Theorem, so named in Section 5.8 of McFadden (1981), who refers to Daly and Zachary (1978) and Williams (1977) as forerunners of the result.

This section presents the SEVI surplus function, compares it with its LEVI counterpart, characterizes how that comparison varies with the dispersion of the scale values, and then provides three applications: compensating variation for a price change, compensating variation for removing an alternative, and parameter identification.

3.1 The Surplus Function and the Option Value of a Choice Set

When $e_j \sim iid$ LEVI, it is well known that the surplus function takes the familiar "log-sum" form:

$$W_J^{\text{LEVI}}(v) = \ln\left[\sum_{j=1}^{J} \exp(v_j)\right] + \gamma = \ln\left[\sum_{j=1}^{J} s_j\right] + \gamma \text{ for } s_j = \exp(v_j).$$

Here, the superscript "LEVI" signifies that the surplus function is based on the assumption that $e_j \sim iid$ LEVI. The next lemma provides the corresponding closed-form expression when $e_j \sim iid$ SEVI.

Lemma 2 *If $e_j \sim iid$* SEVI *across $j = 1, \ldots, J$, then the surplus function, denoted by $W_J^{\text{SEVI}}(v)$, is*

$$\begin{aligned} W_J^{\text{SEVI}}(v) &= \sum_{\ell=1}^{J} (-1)^{\ell} \sum_{|\mathcal{S}|=\ell, \mathcal{S}\subseteq\mathcal{J}} W_\ell^{\text{LEVI}}(-v_{\mathcal{S}}) \\ &= \sum_{\ell=1}^{J} (-1)^{\ell} \sum_{|\mathcal{S}|=\ell, \mathcal{S}\subseteq\mathcal{J}} \ln\left[\sum_{k\in\mathcal{S}} \frac{1}{s_k}\right] - \gamma, \end{aligned} \tag{10}$$

where the inner sum runs over the subsets $\mathcal{S}$ of $\mathcal{J}$ with ℓ elements and $v_{\mathcal{S}} := (v_k)_{k\in\mathcal{S}}$ collects the systematic utilities of the alternatives in $\mathcal{S}$. Furthermore, $W_J^{\text{SEVI}}(v)$ is strictly convex on $\mathbb{R}^{J-1} \times \{0\}$.

Equation (10) shows that $W_J^{\text{SEVI}}(v)$ is an alternating sum of log-sums over all nonempty subsets of alternatives, with the signs of the systematic utilities flipped. While more complex than its LEVI counterpart, this expression remains in closed form and retains strict convexity.

The strict convexity in Lemma 2, which is what the identification argument at the end of this section uses, does not appear in Anderson and de Palma (1999) and follows instead from Theorem 6 of Sørensen and Fosgerau (2022).[7]

To study and compare the surplus functions, we consider a set of $N \geq 1$ alternatives, because we apply the comparison below both to all J alternatives and to subsets of them. It is convenient to work with the recentered surplus functions and to index them by the vector of scale values $s = \exp(v) \in \mathbb{R}^N_{++} := (0, \infty)^N$ rather than by v itself. Accordingly, we denote

$$\widetilde{W}_N^{\text{LEVI}}(s) := W_N^{\text{LEVI}}(\ln s) - \gamma, \qquad \widetilde{W}_N^{\text{SEVI}}(s) := W_N^{\text{SEVI}}(\ln s) + \gamma. \tag{11}$$

[7] The closed-form expression in (10) is due to Anderson and de Palma (1999), who state it in their Section 4.1 and derive it in their Appendix A for cost-minimizing choice with LEVI errors, the setting that is equivalent to utility-maximizing choice with SEVI errors. As with Lemma 1, we record it in the notation used throughout this section and give the derivation in the supplementary appendix.

The tilde thus signals both the recentering and the change of argument. For $\mathcal{M} \in \{\text{LEVI}, \text{SEVI}\}$, the untilded $W_N^{\mathcal{M}}(v)$ retains the systematic utilities as its argument. Recentering a surplus function is equivalent to recentering the random components and thus has no effect on the choice probabilities. It does, however, make the surplus function comparison reflect the shape of the two error distributions rather than their means.

The two recentered surplus functions respond differently to the scale values once $N \geq 3$. The LEVI surplus $\widetilde{W}_N^{\text{LEVI}}(s) = \ln(\sum_{j=1}^N s_j)$ depends on s only through its coordinate sum, whereas the SEVI surplus admits no such reduction. For $N \leq 2$, the two coincide, a degeneracy recorded in Proposition 1(a) below. To describe how the two functions move as the scale values are spread apart, we order vectors of scale values by spread, using the standard notions of majorization and Schur-convexity. For $x, y \in \mathbb{R}_{++}^N$, let $x_{[1]} \geq \cdots \geq x_{[N]}$ denote the coordinates of x in decreasing order. We say that x *majorizes* y, written $x \succ y$, if

$$\sum_{i \leq k} x_{[i]} \geq \sum_{i \leq k} y_{[i]} \text{ for } k = 1, \ldots, N-1, \qquad \text{and} \qquad \sum_{i \leq N} x_{[i]} = \sum_{i \leq N} y_{[i]}.$$

A function $f : \mathbb{R}_{++}^N \to \mathbb{R}$ is *Schur-convex* if $x \succ y$ implies $f(x) \geq f(y)$, and *strictly Schur-convex* if in addition $f(x) > f(y)$ whenever $x \succ y$ and x is not a permutation of y. It is *Schur-concave* (*strictly Schur-concave*) if $-f$ is Schur-convex (strictly Schur-convex). See Marshall et al. (2011, Ch. 1.A and Ch. 3.A) for details on majorization and Schur-convexity.

Majorization is a partial order that ranks vectors by spread: among vectors with the same coordinate sum τ, the least spread vector is $(\tau/N, \ldots, \tau/N)$, and the spread grows as the sum concentrates on fewer coordinates, with $(\tau, 0, \ldots, 0)$ as the limiting case, approached from within $\mathbb{R}_{++}^N$.

For notational convenience, define the function $\phi : \mathbb{R}_{++}^N \to \mathbb{R}$ as

$$\phi(s) := W_N^{\text{SEVI}}(\ln s), \tag{12}$$

which is the same as the surplus function under SEVI but with the vector of scale values $s \in \mathbb{R}_{++}^N$ as its argument.

Theorem 3 *(a) For any $N \geq 1$, ϕ is Schur-convex on $\mathbb{R}_{++}^N$, and ϕ is strictly Schur-convex on $\mathbb{R}_{++}^N$ for $N \geq 3$.*

(b) For any $N \geq 1$, $\widetilde{W}_N^{\text{SEVI}}$ is Schur-convex on $\mathbb{R}_{++}^N$, and $\widetilde{W}_N^{\text{SEVI}}$ is strictly Schur-convex on $\mathbb{R}_{++}^N$ for $N \geq 3$.

The recentered surplus functions are not easy to interpret on their own. The option value is the more meaningful quantity. For a given choice set, the option value is defined as:

$$\text{OV}_N^{\mathcal{M}}(s) := \widetilde{W}_N^{\mathcal{M}}(s) - \max_{j \leq N} \ln s_j = E\left[\max_{j \leq N} \left\{\ln(s_j) + \tilde{e}_j^{\mathcal{M}}\right\}\right] - \max_{j \leq N} E\left(\ln s_j + \tilde{e}_j^{\mathcal{M}}\right), \tag{13}$$

where $\tilde{e}_j^{\mathcal{M}} = e_j^{\mathcal{M}} - E(e_j^{\mathcal{M}})$ is the recentered version of $e_j^{\mathcal{M}}$. The option value measures the gain from choosing after the errors are realized rather than before. It is nonnegative by Jensen's inequality applied to the maximum function, and equals zero when $N = 1$. Since $\max_{j \leq N} \ln s_j$ is common to the two models, comparing option values coincides with comparing recentered surplus functions:

$$\Gamma_N(s) := \text{OV}_N^{\text{LEVI}}(s) - \text{OV}_N^{\text{SEVI}}(s) = \widetilde{W}_N^{\text{LEVI}}(s) - \widetilde{W}_N^{\text{SEVI}}(s). \tag{14}$$

We call $\Gamma_N(s)$ the option-value gap. It is the object through which the rest of this section compares the two models, and it reappears in the compensating variations of Section 3.3 below.

Proposition 1 *Let $s = (s_1, \ldots, s_N) \in \mathbb{R}^N_{++}$ for $N \geq 1$. Then:*

(a) $\Gamma_N(s) = 0$ for $N \leq 2$, and $\Gamma_N(s) > 0$ for $N \geq 3$.

(b) If $s \succ s'$ for $s, s' \in \mathbb{R}^N_{++}$, then $\Gamma_N(s) \leq \Gamma_N(s')$, with strict inequality when $N \geq 3$ and s is not a permutation of s'.

(c) $\Gamma_N(s)$ attains its maximum over $s \in \mathbb{R}^N_{++}$ when $s_1 = \cdots = s_N$.

Proposition 1(a) determines the sign of the option-value gap, while Proposition 1(b) and (c) say how the gap moves as the scale values are spread apart and where it is largest. The first of these shows that, regardless of the scale values, the two models are indistinguishable in option-value terms when $N \leq 2$, and that the LEVI option value is strictly larger when $N \geq 3$. For $N \geq 3$, a LEVI analyst therefore attributes more value to the ability to choose after the errors are realized than a SEVI analyst does, at every configuration of scale values. Figures 3 and 4 in Section 3.3 display this directly: at the full choice set in both figures, the recentered surplus under LEVI lies strictly above its SEVI counterpart, and the vertical segment joining them is the option-value gap $\Gamma_J(s)$.

Proposition 1(b) compares configurations with the same sum of scale values, and that restriction is not an arbitrary normalization. Because $\widetilde{W}^{\text{LEVI}}_N$ depends on s only through $\sum_{j=1}^N s_j$, the configurations sharing a coordinate sum are exactly those to which a LEVI analyst assigns the same surplus. Along any such set, Γ_N varies only through $\widetilde{W}^{\text{SEVI}}_N$, so Part (b) reads the Schur-convexity of Theorem 3(b) as a statement about the gap: spreading the scale values apart raises the SEVI surplus and therefore narrows Γ_N.

Part (c) lifts the restriction to a fixed coordinate sum. Multiplying every scale value by $c > 0$ adds $\ln c$ to both recentered surplus functions and so leaves their difference unchanged, giving $\Gamma_N(cs) = \Gamma_N(s)$ for every $c > 0$. Any configuration can therefore be normalized to the coordinate sum N at no cost, and Part (b) then ranks it against the equal vector $\mathbf{1}_N = (1, \ldots, 1) \in \mathbb{R}^N$, giving $\Gamma_N(s) \leq \Gamma_N(\mathbf{1}_N)$ for all $s \in \mathbb{R}^N_{++}$. The bound is attained when $s_1 = \cdots = s_N$, and for $N \geq 3$ only there.

The results are intuitive. When the scale values are equal, the systematic utilities carry no information and the random components decide the choice entirely. In this case, the shape of the error distribution is fully expressed and the two models are as far apart as they can be. When one scale value dominates, the choice is nearly determined before any error is drawn, the tails almost never bind, and both models return nearly the same surplus. The distributional assumption is priced in proportion to how much work the random component is doing.

Corollary 1(b) and Proposition 1(a) pull in opposite directions, and it is worth seeing why both hold. Decomposing the recentered surplus into the two parts of the utility of the chosen alternative, $\widetilde{W}^{\mathcal{M}}_J(s) = E^{\mathcal{M}}(v_Y|v) + E^{\mathcal{M}}(\tilde{e}_Y|v)$, the corollary says that the first term is larger under SEVI, while the proposition says that the total is larger under LEVI. The two are consistent because the second term is larger under LEVI by more than the first is smaller: the SEVI model chooses the better alternative on the systematic dimension, whereas the LEVI model extracts more from the idiosyncratic one. This is the same thin-upper-tail mechanism as elsewhere in the comparison, now split into its two components.

3.2 Application: Compensating Variation for a Price Change

As the first application of Lemma 2, consider a SEVI model of the form:

$$U_{ij} = V_{ij} + \varepsilon_{ij},\ V_{ij} = \tilde{V}_{ij} + (\mathcal{I}_i - K_j)\,\lambda, \tag{15}$$

where $\varepsilon_{ij} \sim iid$ SEVI for $j = 1, 2, \ldots, J$. Here, we have introduced an additional covariate $\mathcal{I}_i - K_j$ as part of the systematic utility, where $\mathcal{I}_i$ is the income of individual i and K_j is the cost of alternative j. The coefficient $\lambda > 0$ on this covariate represents the marginal utility of income. In this model, the systematic utility that individual i obtains from alternative j consists of two components: the utility $(\mathcal{I}_i - K_j)\lambda$ derived from the expenditure on the numeraire and the utility $\tilde{V}_{ij}$ derived from the characteristics of alternative j. Such a formulation is common in applied welfare analysis using discrete choice models.

Now suppose we change the cost for the m-th alternative from K_m to $K_m + \Delta K_m$ while holding all else fixed. For each individual i, the compensating variation (CV) is defined implicitly via the equation below:

$$\begin{aligned}&\max_{j \in \mathcal{J}} \left\{ \tilde{V}_{ij} + [\mathcal{I}_i + \text{CV} - (K_j + 1\{m = j\} \Delta K_m)]\lambda + \varepsilon_{ij} \right\} \\ &= \max_{j \in \mathcal{J}} \left\{ \tilde{V}_{ij} + (\mathcal{I}_i - K_j)\lambda + \varepsilon_{ij} \right\},\end{aligned}$$

where $1\{\cdot\}$ is the indicator function. That is, CV is the additional income that individual i would need to maintain the same utility when the cost of alternative m changes from K_m to $K_m + \Delta K_m$. Solving for CV yields

$$\text{CV} = \frac{1}{\lambda} \left[\max_j \left\{ \tilde{V}_{ij} + (\mathcal{I}_i - K_j)\lambda + \varepsilon_{ij} \right\} - \max_j \left\{ \tilde{V}_{ij} + [\mathcal{I}_i - (K_j + 1\{m = j\} \Delta K_m)]\lambda + \varepsilon_{ij} \right\} \right].$$

Taking expectations of each maximum and using the definition of the surplus function, the expected value of CV conditional on V and ΔK_m is

$$E(\text{CV}|V, \Delta K_m) = \frac{1}{\lambda} \left[W_J^{\text{SEVI}}(V) - W_J^{\text{SEVI}}(V - \iota_m \Delta K_m \lambda) \right], \tag{16}$$

where ι_m is the unit vector with a value of 1 at the m-th element and 0 elsewhere. Here the individual subscript is suppressed, V being treated as the random vector of systematic utilities of an individual drawn at random from the population. We can average $E(\text{CV}|V, \Delta K_m)$ over the distribution of V to get the expected compensating variation over the population:

$$E(\text{CV}|\Delta K_m) := E[E(\text{CV}|V, \Delta K_m)|\Delta K_m].$$

In empirical applications, we can estimate $E(\text{CV}|V, \Delta K_m)$ and $E(\text{CV}|\Delta K_m)$ using sample averages after plugging in parameter estimates.

3.3 Application: Compensating Variation for Removing an Alternative

As a second and closely related application, we can employ the surplus function to define the compensating variation when an alternative is removed from the choice set. Using the same model as in (15), we investigate the minimum compensation necessary for individual i so that they would not be made worse off if alternative k were eliminated from their choice set. Denote this compensation as $\text{CV}(k)$. Then, $\text{CV}(k)$ solves

$$\max_{j \in \mathcal{J} \setminus \{k\}} \left\{ \tilde{V}_{ij} + (\mathcal{I}_i + \text{CV}(k) - K_j)\lambda + \varepsilon_{ij} \right\} = \max_{j \in \mathcal{J}} \left\{ \tilde{V}_{ij} + (\mathcal{I}_i - K_j)\lambda + \varepsilon_{ij} \right\}.$$

By a similar argument to that used to derive (16), we obtain

$$\text{CV}(k) = \frac{1}{\lambda} \left[\max_{j \in \mathcal{J}} \left\{ \tilde{V}_{ij} + (\mathcal{I}_i - K_j)\lambda + \varepsilon_{ij} \right\} - \max_{j \in \mathcal{J} \setminus \{k\}} \left\{ \tilde{V}_{ij} + (\mathcal{I}_i - K_j)\lambda + \varepsilon_{ij} \right\} \right].$$

Taking expectations of each maximum and using the definition of the surplus function,

$$\begin{aligned} E\left[\mathrm{CV}\left(k\right)|V\right] &= \frac{1}{\lambda}\left[W_J^{\mathrm{SEVI}}(V) - W_{J-1}^{\mathrm{SEVI}}(V_{-k})\right] \\ &= \frac{1}{\lambda}\left[\widetilde{W}_J^{\mathrm{SEVI}}(\exp(V)) - \widetilde{W}_{J-1}^{\mathrm{SEVI}}(\exp(V_{-k}))\right]. \end{aligned}$$

Since the conditional compensating variation given $V = v$ depends on v only through $s = \exp(v)$, we hereafter index it by the scale values and write $\mathrm{CV}^{\mathrm{SEVI}}(k, s) := E\left[\mathrm{CV}\left(k\right)|V = v\right]$. We can define $\mathrm{CV}^{\mathrm{LEVI}}(k, s)$ similarly under the LEVI model. Clearly, $\mathrm{CV}^{\mathrm{LEVI}}(k, s) = \mathrm{CV}^{\mathrm{SEVI}}(k, s)$ for $J = 2$.

The next proposition gives closed-form expressions for the two compensating variations at every $J \geq 2$. Take the removed alternative to be alternative 1, without loss of generality, and normalize the marginal utility of income to $\lambda = 1$ in what follows.

Proposition 2 *Let $s = (s_1, \ldots, s_J) \in \mathbb{R}^J_{++}$ be the scale values for $J \geq 2$ alternatives, and let the marginal utility of income be $\lambda = 1$. Then:*

(a) For the LEVI model,

$$\mathrm{CV}^{\mathrm{LEVI}}(1, s) = \ln\left(1 + \frac{s_1}{\sum_{k\in\mathcal{J}_{-1}} s_k}\right). \tag{17}$$

(b) For the SEVI model,

$$\mathrm{CV}^{\mathrm{SEVI}}(1, s) = \sum_{\ell=1}^{J-1}(-1)^{\ell+1} \sum_{|\mathcal{S}|=\ell, \mathcal{S}\subseteq\mathcal{J}_{-1}} \ln\left(1 + s_1 \sum_{j\in\mathcal{S}} s_j^{-1}\right). \tag{18}$$

Both compensating variations depend on s only through the ratios s_1/s_k, $k \geq 2$. To compare them, define

$$D_{\mathrm{CV}}(1, s) := \mathrm{CV}^{\mathrm{SEVI}}(1, s) - \mathrm{CV}^{\mathrm{LEVI}}(1, s) \tag{19}$$

as the difference between the two compensating variations. Write $s_{-1} = (s_2, \ldots, s_J)$ for the rival scale values and $|s_{-1}| := \sum_{j\in\mathcal{J}_{-1}} s_j$ for their sum. Holding s_{-1} fixed, define

$$D^{\infty}_{\mathrm{CV}}\left(s_{-1}\right) := \lim_{s_1\to\infty} D_{\mathrm{CV}}(1, s) \tag{20}$$

as the limit of $D_{\mathrm{CV}}(1, s)$ when the removed alternative becomes dominant (i.e., $s_1 \to \infty$). Proposition 3(d)(i) below shows that this limit exists and identifies it. We suppress the dependence of $D^{\infty}_{\mathrm{CV}}\left(s_{-1}\right)$ on s_{-1} and write D^{∞}_{CV} when s_{-1} is held fixed and no confusion arises. Note that D_{CV} is oriented as SEVI minus LEVI, the reverse of Γ_N. The option-value gap is oriented so as to be nonnegative, whereas the comparisons of choice probabilities and compensating variations are stated from the SEVI analyst's viewpoint. Thus $\Gamma_N > 0$ and $D_{\mathrm{CV}} < 0$ both say that the LEVI model assigns the larger value.

With $\lambda = 1$, the recentering constant cancels between the two surplus functions, so $\mathrm{CV}^{\mathcal{M}}(1, s) = \widetilde{W}^{\mathcal{M}}_J(s) - \widetilde{W}^{\mathcal{M}}_{J-1}(s_{-1})$. Plugging this into (19) and using (14) yields, for every $J \geq 2$ and every s,

$$D_{\mathrm{CV}}(1, s) = \Gamma_{J-1}\left(s_{-1}\right) - \Gamma_J\left(s\right). \tag{21}$$

The comparison of compensating variations is therefore a comparison of two option-value gaps: that of the alternatives that remain, less that of the full choice set. The next proposition signs $D_{\mathrm{CV}}(1, s)$.

Proposition 3 *Let $s = (s_1, \ldots, s_J) \in \mathbb{R}^J_{++}$ be the scale values for $J \geq 2$ alternatives, and let $\lambda = 1$. Then:*

(a) Let $J = 3$. Then $D_{\mathrm{CV}}(1, s) < 0$ for every $s \in \mathbb{R}^3_{++}$.

(b) Let $J \geq 3$. If $s_1 \leq \min_{j \neq 1} s_j$, then $D_{\mathrm{CV}}(1, s) < 0$.

(c) As $s_1 \to 0$ while holding $s_2, \ldots, s_J$ fixed,

$$\mathrm{CV}^{\mathrm{LEVI}}(1, s) = \left(\sum_{k=2}^{J} s_k \right)^{-1} s_1 \left(1 + o\left(1\right)\right),$$

$$\mathrm{CV}^{\mathrm{SEVI}}(1, s) = (J-2)! \left(\prod_{k=2}^{J} s_k^{-1} \right) s_1^{J-1} \left(1 + o\left(1\right)\right). \tag{22}$$

(d)(i) The limit (20) *exists and equals $\Gamma_{J-1}(s_{-1})$, which is nonnegative, and strictly positive for $J \geq 4$. (ii) $D_{\mathrm{CV}}(1, s) \leq \Gamma_{J-1}(s_{-1})$, with strict inequality for $J \geq 3$. (iii) If $s_1 \geq \max_{j \neq 1} s_j$, then $D_{\mathrm{CV}}(1, s)$ is nondecreasing in s_1 for fixed s_{-1}, strictly so for $J \geq 3$.*

For $J = 3$, Proposition 3(a) shows that $\mathrm{CV}^{\mathrm{SEVI}}(1, s) < \mathrm{CV}^{\mathrm{LEVI}}(1, s)$ regardless of how the systematic utilities compare.

Proposition 3(b) extends this to every $J \geq 3$ on the *laggard cone* $\left\{ s \in \mathbb{R}^J_{++} : s_1 \leq \min_{j \neq 1} s_j \right\}$, on which alternative 1 is the least favorable option: the welfare loss from excluding it is then strictly smaller under SEVI than under LEVI. Unlike Proposition 3(c), this is a non-asymptotic statement, valid on the whole cone. The proof of Proposition 3(b) is in the supplementary appendix. It uses Theorem 1 together with Lemma S.5, which connects the surplus function to the choice probabilities. Figure 3 displays Proposition 3(b) as a difference in differences. Each model is evaluated at the full choice set and at the subset with alternative 1 removed. The drop in the surplus function for a given model is its compensating variation, and the gap between the two surplus functions narrows upon the removal of a laggard alternative. The magnitude of the contraction is $|D_{\mathrm{CV}}(1, s)|$, drawn against the dotted path along which the SEVI surplus would have fallen had it dropped by the LEVI amount.

Proposition 3(c) sharpens the same comparison quantitatively in the limit. Let $J \geq 3$. Then $\mathrm{CV}^{\mathrm{LEVI}}(1, s)$ vanishes at rate s_1 as $s_1 \to 0$, whereas $\mathrm{CV}^{\mathrm{SEVI}}(1, s)$ vanishes at the much faster rate s_1^{J-1}, so that $\mathrm{CV}^{\mathrm{SEVI}}(1, s)/\mathrm{CV}^{\mathrm{LEVI}}(1, s) \to 0$. The welfare loss from removing alternative 1 vanishes under both specifications, so the content of the comparison is relative rather than absolute: not only is the SEVI loss smaller throughout the laggard cone, but it is also smaller by a factor of order s_1^{J-2} as the alternative becomes negligible. The SEVI model thus discounts the laggard alternative far more heavily than the LEVI model does. The same pair of rates governs the choice probabilities themselves: by Theorem 1(c), $P^{\mathrm{LEVI}}(Y = 1|v)$ is of order s_1 and $P^{\mathrm{SEVI}}(Y = 1|v)$ of order s_1^{J-1} as $s_1 \to 0$.

Proposition 3(d) concerns the opposite extreme, in which the removed alternative becomes dominant. Proposition 3(d)(i) identifies the limit as $\Gamma_{J-1}(s_{-1})$, so, by the scale invariance of Γ_{J-1}, the limit depends on s_{-1} only through the ratios s_j/s_k, $j, k \geq 2$. By (21), the difference $D^{\infty}_{\mathrm{CV}}(s_{-1}) - D_{\mathrm{CV}}(1, s)$ at any finite s_1 is exactly $\Gamma_J(s)$, the option-value gap of the full choice set, which is nonnegative and vanishes as $s_1 \to \infty$. The limit $D^{\infty}_{\mathrm{CV}}(s_{-1})$ is therefore approached by $D_{\mathrm{CV}}(1, s)$ strictly from below whenever $J \geq 3$. Figure 4 shows the same construction as Figure 3 with the sign reversed: the configuration drawn has $J \geq 4$, which Proposition 3(a) makes necessary for D_{CV} to be positive, and s_1 large enough that the gap between the models *widens* on removal. The widening equals $D_{\mathrm{CV}}(1, s)$, which is positive.

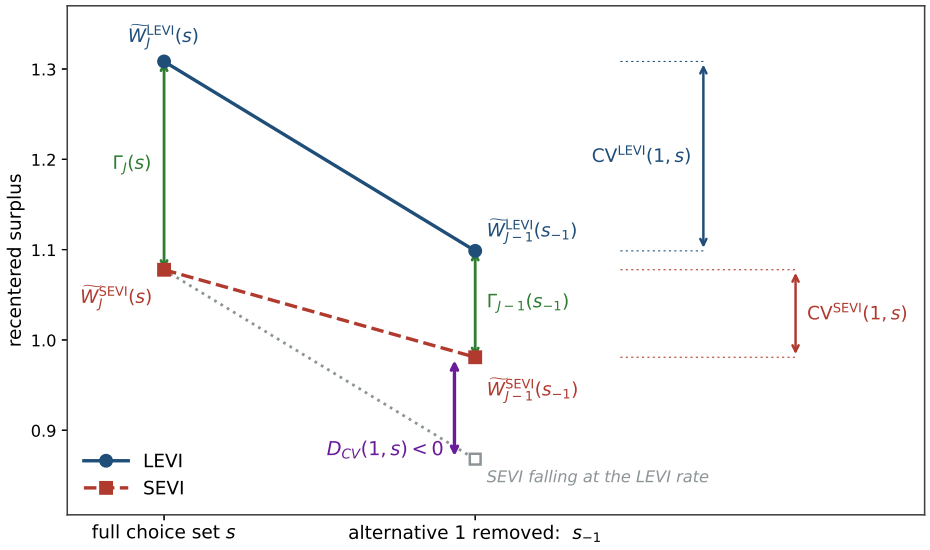


Figure 3: Removing a systematic laggard ($s_1 \leq \min_{j \neq 1} s_j$): the recentered surplus falls less under SEVI than under LEVI, so $D_{\text{CV}}(1,s) < 0$

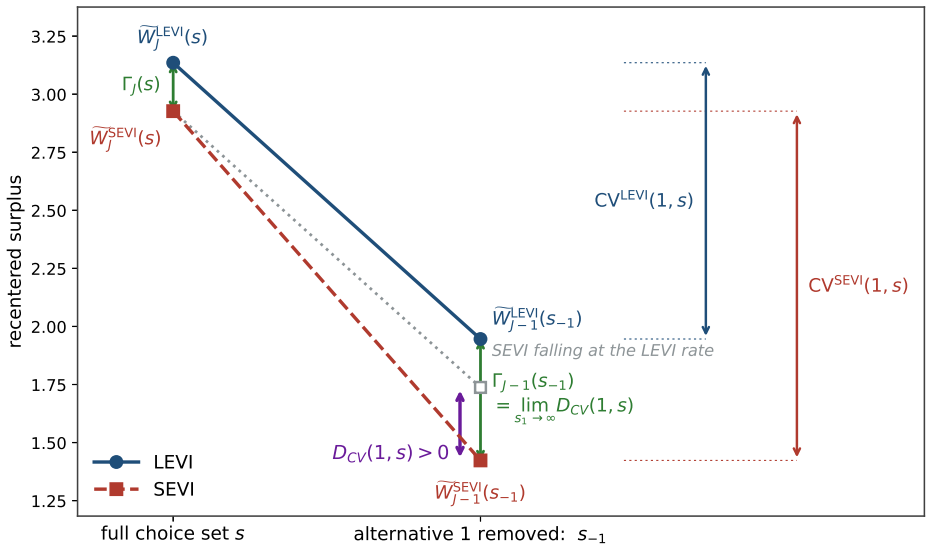


Figure 4: Removing a dominant alternative (s_1 large): the recentered surplus falls more under SEVI than under LEVI, so $D_{\text{CV}}(1,s) > 0$

Proposition 3(d)(iii) adds monotonicity along the way. On the *leader region* $s_1 \geq \max_{j \neq 1} s_j$, raising s_1 with s_{-1} held fixed spreads the normalized scale values further apart, so the option-value gap Γ_J of the full choice set falls by Proposition 1(b), and D_{CV} rises by (21). It therefore climbs monotonically to $D_{\mathrm{CV}}^{\infty}(s_{-1})$, which Part (d)(i) makes strictly positive for $J \geq 4$. Hence D_{CV} is positive once s_1 is large enough: removing a sufficiently strong systematic leader costs more under the SEVI model than under the LEVI model. The threshold lies inside the leader region rather than at its edge, so membership alone does not suffice. At $s = (1, 1, 1, 1)$, alternative 1 is weakly the largest, yet $D_{\mathrm{CV}}(1, s) < 0$, and it is still negative at $s = (1.2, 1, 1, 1)$, where alternative 1 is a strict leader. With $J - 1$ rivals of common scale 1, the laggard cone and the leader region together exhaust the range of s_1, and the difference changes sign exactly once, when s_1 exceeds roughly 6.9, 4.4, 3.6, 3.2, and 2.9 for $J = 4, \ldots, 8$.

The upper bound $\Gamma_{J-1}(s_{-1})$ in Proposition 3(d)(ii) depends on both J and s_{-1}. It can be relaxed to a bound that depends on J alone. Applying Proposition 1(c) with $N = J - 1$, we have, for every $s \in \mathbb{R}^J_{++}$,

$$D_{\mathrm{CV}}(1, s) \leq \Gamma_{J-1}(\mathbf{1}_{J-1}) = \ln(J-1) + \sum_{\ell=1}^{J-1} (-1)^{\ell+1} \binom{J-1}{\ell} \ln(\ell),$$

where $\mathbf{1}_{J-1}$ is the vector of ones and the closed form follows from Lemma 2 and (14) at equal scale values. The bound depends on the choice set only through its size, and it cannot be improved: it is approached at equal rivals as $s_1 \to \infty$.

The laggard and dominant cases have a common explanation in the shapes of the two error distributions. The LEVI distribution is more dispersed at the upper end, so it assigns a higher option value to every choice set: a large random shock is more likely to override a low systematic utility. The thinner SEVI right tail keeps choices closer to systematic utilities. A systematic laggard is therefore worth little under SEVI beyond its systematic utility, and removing it costs less than under LEVI. A dominant alternative is the opposite case: the SEVI model has concentrated choice on it, whereas the LEVI model has already spread choice across the rivals, so removing it costs more under SEVI once $J \geq 4$.

Between the laggard cone and the dominance limit, an analytical comparison is difficult, and the ranking is not fixed by any of the results above. Section S.H of the supplementary appendix computes the ratio of the two compensating variations numerically and shows that it may cross 1 when the excluded alternative is neither a systematic laggard nor dominant.

3.4 Application: Identification

As a third application, we use Lemma 2 to establish the identification of the SEVI model. Note that the model under the vector of systematic utilities $V = (V_1, \ldots, V_J)$ is equivalent to that under $V^o = (V_1 - V_J, \ldots, V_{J-1} - V_J, 0)$, as the choice problem depends only on utility differences. For the purpose of studying identification, we normalize the systematic utility of the last alternative to zero and maintain this normalization in the rest of this subsection. Each vector of systematic utilities $V \in \mathbb{R}^J_o := \mathbb{R}^{J-1} \times \{0\}$ maps into a vector of choice probabilities $P \in \mathbb{S}^{J-1} := \{P = (P_1, \ldots, P_{J-1}) : P_j > 0 \text{ and } \sum_{j=1}^{J-1} P_j < 1\}$, the strict inequalities holding because the random components have full support. The question is whether we can deduce the systematic utilities from the choice probabilities.

Define the conjugate surplus function:

$$W_*^{\text{SEVI}}(P) := \sup_{v \in \mathbb{R}_o^J} \left[\sum_{j=1}^{J-1} v_j P_j - W_J^{\text{SEVI}}(v) \right] \text{ for any } P \in \mathbb{S}^{J-1}.$$

Since every $P \in \mathbb{S}^{J-1}$ is interior to the probability simplex, the supremum is finite, so the conjugate surplus function is well-defined; moreover, the strict convexity of $W_J^{\text{SEVI}}(v)$ established in Lemma 2 makes the maximizing v unique. Since $W_J^{\text{SEVI}}(v)$ is known, it follows by construction that all of $W_*^{\text{SEVI}}(\cdot) : \mathbb{S}^{J-1} \to \mathbb{R}$ and $\left\{\partial W_*^{\text{SEVI}}(\cdot)/\partial P_j : \mathbb{S}^{J-1} \to \mathbb{R}\right\}_{j=1}^{J-1}$ are known functions. Using Theorem 7 of Sørensen and Fosgerau (2022), we have:

$$V_j = \left. \frac{\partial W_*^{\text{SEVI}}(P)}{\partial P_j} \right|_{P=[P^{\text{SEVI}}(Y=1|V),\ldots,P^{\text{SEVI}}(Y=J-1|V)]}$$

for $j = 1, \ldots, J-1$.[8] That is, we can recover the vector of systematic utilities by plugging the vector of choice probabilities into $\left\{\partial W_*^{\text{SEVI}}(\cdot)/\partial P_j\right\}_{j=1}^{J-1}$.

Since the vector of empirical market shares converges to the vector of choice probabilities,[9] the vector of choice probabilities is identified, and hence the vector of systematic utilities is also identified. The problem of identifying the model parameter β_0 in a SEVI model thus reduces to the problem of identifying β_0 from the (joint) distribution of $V^*(\beta_0) := (V_1(\beta_0) - V_J(\beta_0), \ldots, V_{J-1}(\beta_0) - V_J(\beta_0)) \in \mathbb{R}^{J-1}$. Intuitively, if we have an infinite number of independent draws from the distribution of $V^*(\beta_0)$, can we pin down β_0? Mathematically, if

$$\Pr(V^*(\beta_1) \neq V^*(\beta_2)) > 0 \text{ for any } \beta_1 \neq \beta_2, \tag{23}$$

then β_0 is identified. The condition is also necessary: if it fails, there are $\beta_1 \neq \beta_2$ with $V^*(\beta_1) = V^*(\beta_2)$ almost surely, and the two parameters generate the same choice probabilities. This identification condition requires that, for two different model parameters, there is a positive probability that the resulting vectors of (normalized) systematic utilities are different. For example, in the linear case where $V_i^*(\beta) = [(X_{i1} - X_{iJ})\beta, \ldots, (X_{i,J-1} - X_{iJ})\beta]$, it rules out the following case: there exists $\delta \neq 0$ satisfying $(X_{ij} - X_{iJ})\delta = 0$ almost surely for every $j \in \{1, \ldots, J-1\}$. The identification condition therefore holds whenever the variables in $X_{ij} - X_{iJ}$ are free of perfect multicollinearity for at least one j.

The identification condition in (23) does not involve the distribution of ε_{ij}. Therefore, it is applicable to any RUM, provided that the unobserved random components $(\varepsilon_{i1}, \ldots, \varepsilon_{iJ})$ are independent of the systematic utilities $(V_{i1}, \ldots, V_{iJ})$ and are absolutely continuous with full support $\mathbb{R}^J$. See Corollary 1 of Sørensen and Fosgerau (2022) for further discussion. In particular, it also characterizes identification in a LEVI model. Hence, a SEVI model is identified if and only if the corresponding LEVI model is identified, and no additional identification challenges arise under SEVI.

[8] This may be of independent interest in IO/marketing applications if the objective is to derive the systematic utilities from the observed market shares, but here we focus on identification.

[9] If V has a continuous distribution, then for each $v \in \mathbb{R}_o^J$, the associated market shares can be computed as the empirical choice frequencies of all observations lying in an appropriate shrinking neighborhood of v.

4 The MLE and QMLE of a SEVI Model

4.1 The Estimators

Consider the following conditional SEVI model

$$U_{ij}(\beta_0) = V_{ij}(\beta_0) + \varepsilon_{ij} \text{ and } V_{ij}(\beta_0) = X_{ij}\beta_0$$

for $j = 1, \ldots, J$, where $X_{ij} \in \mathbb{R}^{1\times L}$ and $\varepsilon_{ij} \sim iid$ SEVI.

The SEVI framework supports the same linear parametrizations as the standard LEVI model: the conditional model given above, which uses alternative-specific attributes X_{ij}; the multinomial model, which uses individual characteristics Z_i with alternative-specific coefficients β_j; and hybrids of the two. The latter two reduce to the conditional form. Interacting a covariate with dummies for $J-1$ of the alternatives and placing the products in X_{ij} gives that covariate an alternative-specific coefficient; taking the covariate to be a constant gives alternative-specific constants. Dropping the J-th dummy is exactly the normalization required by the multinomial model (e.g., $\beta_J = 0$), as in the standard LEVI model. Given the reduction to the conditional form, we focus on the conditional SEVI without loss of generality.

Given a simple random sample $\{Y_i, X_{i1}, \ldots, X_{iJ}\}_{i=1}^n$ from the above model, we consider estimating β_0 by MLE and QMLE. The MLE is the maximum likelihood estimator when the likelihood function is based on a correctly specified SEVI model, and the QMLE is the maximum likelihood estimator based on an incorrectly specified LEVI model.

The negative log-likelihood function of the model under the correct SEVI specification is

$$\ell_{\text{SEVI}}(\beta) = \sum_{i=1}^n \ell_i^{\text{SEVI}}(\beta) \text{ for } \ell_i^{\text{SEVI}}(\beta) = -\sum_{k=1}^J Y_{ik} \log\left(P_{ik}^{\text{SEVI}}(\beta)\right),$$

where $Y_{ik} := 1\{Y_i = k\}$ and $P_{ik}^{\text{SEVI}}(\beta) := P^{\text{SEVI}}(Y_i = k|v_i(\beta))$ is defined as in (3) with v_i replaced by $v_i(\beta)$. By minimizing the above negative log-likelihood function $\ell_{\text{SEVI}}(\beta)$, we obtain the MLE:

$$\hat{\beta}_{\text{MLE}} = \arg\min_{\beta\in\mathcal{B}} \ell_{\text{SEVI}}(\beta),$$

where $\mathcal{B}$ is a large enough compact set in $\mathbb{R}^L$.

To compute the MLE, we can supply the following closed-form gradient to a gradient-based algorithm:

$$\frac{\partial \ell_{\text{SEVI}}(\beta)}{\partial \beta} = \sum_{i=1}^n \frac{\partial \ell_i^{\text{SEVI}}(\beta)}{\partial \beta} := \sum_{i=1}^n S_i^{\text{SEVI}}(\beta),$$

where

$$S_i^{\text{SEVI}}(\beta) = \sum_{j=1}^J \frac{\partial \ell_i^{\text{SEVI}}(\beta)}{\partial v_{ij}(\beta)} \frac{\partial v_{ij}(\beta)}{\partial \beta} = -\sum_{j=1}^J \sum_{k=1}^J \frac{Y_{ik}}{P_{ik}^{\text{SEVI}}(\beta)} \frac{\partial P_{ik}^{\text{SEVI}}(\beta)}{\partial v_{ij}(\beta)} X_{ij}'.$$

In the above, the derivative $\frac{\partial P_{ik}^{\text{SEVI}}(\beta)}{\partial v_{ij}(\beta)}$ is given in (8) or (9) after plugging in the parameter β.

The next theorem establishes the consistency of the MLE.

Theorem 4 *Let $\{Y_i, X_{i1}, \ldots, X_{iJ}\}_{i=1}^n$ be a simple random sample from the SEVI model above. Assume that (i) $\mathcal{B} \subset \mathbb{R}^L$ is compact and $\beta_0 \in \mathcal{B}$; (ii) $E(\max_{j\in\mathcal{J}} \|X_{ij}\|) < \infty$; and (iii) the identification condition (23) holds. Then $\hat{\beta}_{\text{MLE}} \to^p \beta_0$.*

Assumption (i) is standard, and Assumption (ii) is needed to prove uniform convergence. Lemma S.6 in the supplementary appendix bounds the per-observation negative log-likelihood uniformly over $\mathcal{B}$ by

$$0 \leq \sup_{\beta \in \mathcal{B}} \ell_i^{\text{SEVI}}(\beta) \leq 2(J-1) \max_{k \in \mathcal{J}} \|X_{ik}\| \sup_{\beta \in \mathcal{B}} \|\beta\| + \log \frac{J^J}{(J-1)!}. \tag{24}$$

Correspondingly, we can show the bounds under LEVI:

$$0 \leq \sup_{\beta \in \mathcal{B}} \ell_i^{\text{LEVI}}(\beta) \leq 2 \max_{k \in \mathcal{J}} \|X_{ik}\| \sup_{\beta \in \mathcal{B}} \|\beta\| + \log J.$$

The two upper bounds differ only in the coefficient on the moment term, 2 under LEVI against $2(J-1)$ under SEVI, and in the additive constant. In both cases, a finite first moment of the covariates suffices for a uniform law of large numbers, and the combinatorial structure of (4) imposes no additional moment requirement.

The factor $J-1$ in the first term of the upper bound in (24) is the one identified in Theorem 1(c). By (7), $-\log P^{\text{SEVI}}(Y=1|v)$ grows like $(J-1)|v_1|$ as $v_1 \to -\infty$, against $|v_1|$ under LEVI, so an observation in which a systematic laggard is chosen contributes asymptotically $J-1$ times as much to the negative log-likelihood as it would under LEVI. The envelope and the laggard tail are therefore one fact stated twice.

Assumption (iii) is the identification condition of Section 3.4, imposed without modification. It transfers to the likelihood because that section does more than identify the model: the conjugate surplus function delivers $V_j = \partial W_*^{\text{SEVI}}(P)/\partial P_j$, an explicit inverse, so the map $v \mapsto P^{\text{SEVI}}(\cdot|v)$ is one-to-one on $\mathbb{R}_o^J$. Two parameter values that generate the same choice probabilities therefore generate the same normalized systematic utilities, and (23) then forces the two values to coincide. This is what makes β_0 the unique minimizer of the population criterion. Note that Assumption (iii) is not specific to the SEVI model either: as Section 3.4 shows, (23) does not involve the distribution of the random component at all.

Given the score $S_i^{\text{SEVI}}(\beta)$ and the consistent estimator $\hat{\beta}_{\text{MLE}}$, we can estimate the asymptotic variance of $\sqrt{n}(\hat{\beta}_{\text{MLE}} - \beta_0)$ by

$$\hat{\Omega} = \left[\frac{1}{n} \sum_{i=1}^{n} S_i^{\text{SEVI}}(\hat{\beta}_{\text{MLE}}) S_i^{\text{SEVI}}(\hat{\beta}_{\text{MLE}})' \right]^{-1}.$$

Under some additional regularity conditions, including that β_0 is an interior point of $\mathcal{B}$ and that $E\left[S_i^{\text{SEVI}}(\beta_0) S_i^{\text{SEVI}}(\beta_0)'\right]$ is nonsingular, we have

$$\hat{\Omega}^{-1/2} \sqrt{n}(\hat{\beta}_{\text{MLE}} - \beta_0) \to^d \mathcal{N}(0, I_L).$$

Once consistency is established, the arguments for the above are standard, so we omit them and refer the reader to Chapter 13 of Wooldridge (2002).

If we want to test $H_0 : R\beta_0 = r$ against $H_1 : R\beta_0 \neq r$ for some $q \times L$ matrix R with row rank q and $q \times 1$ vector r, we can form the Wald statistic and show that it is asymptotically chi-squared under the null:

$$n\left(R\hat{\beta}_{\text{MLE}} - r\right)' \left(R\hat{\Omega}R'\right)^{-1} \left(R\hat{\beta}_{\text{MLE}} - r\right) \to^d \chi_q^2.$$

In particular, we can use the square roots of the diagonal elements of

$$\left[\sum_{i=1}^{n} S_i^{\text{SEVI}}(\hat{\beta}_{\text{MLE}}) S_i^{\text{SEVI}}(\hat{\beta}_{\text{MLE}})'\right]^{-1},$$

denoted by $\hat{\sigma}_\ell$ for $\ell = 1, \ldots, L$, as the standard errors for the elements of $\hat{\beta}_{\text{MLE}}$. We can construct a 95% confidence interval as $[\hat{\beta}_{\text{MLE},\ell} - 1.96\hat{\sigma}_\ell, \hat{\beta}_{\text{MLE},\ell} + 1.96\hat{\sigma}_\ell]$.

If the random components follow SEVI, but we employ the maximum likelihood method under the incorrect assumption that they follow LEVI, we obtain the QMLE:[10]

$$\hat{\beta}_{\text{QMLE}} = \arg\min_{\beta \in \mathcal{B}} \ell_{\text{LEVI}}(\beta) \text{ where } \ell_{\text{LEVI}}(\beta) = \sum_{i=1}^{n} \ell_i^{\text{LEVI}}(\beta) \text{ for} \tag{25}$$

$$\ell_i^{\text{LEVI}}(\beta) = -\sum_{k=1}^{J} Y_{ik} \log\left(P_{ik}^{\text{LEVI}}(\beta)\right) = -\sum_{k=1}^{J} Y_{ik}\left[X_{ik}\beta\right] + \log\left\{\sum_{j=1}^{J} \exp\left[X_{ij}\beta\right]\right\}.$$

The asymptotic properties of the QMLE are also standard. In particular, we use the sandwich form for estimating the asymptotic variance of $\hat{\beta}_{\text{QMLE}}$ when the LEVI model is not assumed to be correctly specified.

4.2 On the Estimation of Parameter Ratios

By Theorem 4, $\hat{\beta}_{\text{MLE}}$ is consistent for β_0, but in general, $\hat{\beta}_{\text{QMLE}}$ is not. However, the ratio of the individual elements of $\hat{\beta}_{\text{QMLE}}$, say $\hat{\beta}_{\text{QMLE},2}/\hat{\beta}_{\text{QMLE},1}$, may still be consistent for the true ratio, namely $\beta_{0,2}/\beta_{0,1}$, for a nonzero $\beta_{0,1}$. In this subsection, we provide sufficient conditions for the consistency of the ratio estimator, even though the distribution of the random utility component is misspecified.

We consider an RUM with $V_{ij}(\beta_0) = X_{ij}\beta_0$ and *specify* $\varepsilon_{ij} \sim iid$ Q for some distribution Q. This includes LEVI as an example, but Q may be more general and incorrectly specified. Denote the choice probability derived under Q by $P^{\text{Q}}(Y_i = k|V_i(\beta))$. Let β_{Q}^* be the pseudo-true value defined by

$$\beta_{\text{Q}}^* = \arg\min_{\beta \in \mathcal{B}} -E\sum_{k=1}^{J} Y_{ik} \log P^{\text{Q}}(Y_i = k|V_i(\beta)),$$

the population counterpart of $\hat{\beta}_{\text{Q}} := \arg\min_{\beta \in \mathcal{B}} -n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{J} Y_{ik} \log P^{\text{Q}}(Y_i = k|V_i(\beta))$.

We assume that β_{Q}^* is an interior point of the compact set $\mathcal{B}$ so that β_{Q}^* satisfies the first-order conditions. The derivative of the population objective function with respect to β is

$$-E\sum_{k=1}^{J} Y_{ik} \frac{\partial \log P^{\text{Q}}(Y_i = k|V_i(\beta))}{\partial \beta} = -EX_i' \sum_{k=1}^{J} Y_{ik} \lambda_{ik}^{\text{Q}}(X_i\beta) = -EX_i' \Lambda_i^{\text{Q}}(X_i\beta) \overrightarrow{Y}_i,$$

[10] Here, we focus on the QMLE based on the LEVI likelihood. Other QMLEs may be considered. For instance, we might explore the QMLE derived from the specification that the random utility components follow a generalized extreme value distribution. As long as the marginal distribution is incorrectly specified, such an alternative QMLE will generally be inconsistent for the preference parameters.

where

$$X_i = (X_{i1}', \ldots, X_{iJ}')' \in \mathbb{R}^{J \times L},$$
$$\lambda_{ik}^{\mathrm{Q}}(X_i\beta) = \left.\frac{\partial \log P^{\mathrm{Q}}(Y_i = k | V_i = v_i)}{\partial v_i}\right|_{v_i = X_i\beta} \in \mathbb{R}^{J \times 1},$$
$$\Lambda_i^{\mathrm{Q}}(X_i\beta) = \left[\lambda_{i1}^{\mathrm{Q}}(X_i\beta), \ldots, \lambda_{iJ}^{\mathrm{Q}}(X_i\beta)\right] \in \mathbb{R}^{J \times J}, \ \overrightarrow{Y}_i = (Y_{i1}, \ldots, Y_{iJ})' \in \mathbb{R}^{J \times 1}.$$

So, β_{Q}^* satisfies $E\left[X_i'\Lambda_i^{\mathrm{Q}}(X_i\beta_{\mathrm{Q}}^*)\overrightarrow{Y}_i\right] = \mathbf{0} \in \mathbb{R}^L$.

Theorem 5 *Assume that (i)* $\hat{\beta}_{\mathrm{Q}} \to^p \beta_{\mathrm{Q}}^* \in \mathrm{Int}(\mathcal{B})$*, and* β_{Q}^* *is the unique solution to*

$$E\left[X_i'\Lambda_i^{\mathrm{Q}}(X_i\beta)\overrightarrow{Y}_i\right] = \mathbf{0} \in \mathbb{R}^L;$$

(ii) $\beta_0 \neq 0$*, and there exists a nonzero scalar* δ_1° *such that* $E\left[(X_i\beta_0)'\Lambda_i^{\mathrm{Q}}(X_i\beta_0\delta_1^\circ)\overrightarrow{Y}_i\right] = 0$*;*

(iii) $E[X_{ij,\ell}|X_{i1}\beta_0, X_{i2}\beta_0, \ldots, X_{iJ}\beta_0] = \Theta_{0,\ell} + (X_{ij}\beta_0)\Theta_{1,\ell}$ *for some scalars* $\Theta_{0,\ell}$ *and* $\Theta_{1,\ell}$ *and for* $\ell = 1, \ldots, L$*.*

Then $\beta_{\mathrm{Q}}^* = \delta_1^\circ\beta_0$*.*

Theorem 5 shows that while the QMLE $\hat{\beta}_{\mathrm{Q}}$ may not be consistent for β_0, it converges to β_0 up to a multiplicative factor. In particular, if $\beta_{0,m_1} \neq 0$, then

$$\frac{\hat{\beta}_{\mathrm{Q},m_2}}{\hat{\beta}_{\mathrm{Q},m_1}} \to^p \frac{\beta_{0,m_2}}{\beta_{0,m_1}}.$$

This says that the plug-in ratio estimator is consistent for the true ratio.

The equation $E\left[(X_i\beta_0)'\Lambda_i^{\mathrm{Q}}(X_i\beta_0\delta_1^\circ)\overrightarrow{Y}_i\right] = 0$ in Assumption (ii) of Theorem 5 is a single equation with one scalar unknown $\delta_1^\circ \in \mathbb{R}$, so the assumption is a relatively mild one. It is much weaker than assuming the existence of a scalar δ_1° that solves the system of L equations $E\left[X_i'\Lambda_i^{\mathrm{Q}}(X_i\beta_0\delta_1^\circ)\overrightarrow{Y}_i\right] = \mathbf{0}$. Assumption (iii) of Theorem 5 holds if $X_{i1}, \ldots, X_{iJ}$ are independent and each follows a multivariate normal or elliptical distribution. This condition is sufficient but not necessary. There may exist other distributions of $X_{i1}, \ldots, X_{iJ}$ or specific ways to relax the independence assumption under which the proportionality result holds. In empirical applications, the ratio estimator could still be close to the true ratio even though Assumption (iii) does not hold.

Theorem 5 is proved along the lines of the proof of Ruud's (1986) Theorem 1, which does not cover multinomial choice models. Some adjustments to that proof are needed to account for two main differences: there is no constant term in a multinomial choice model, and the covariate X_i in this model is a matrix rather than a vector.

4.3 The Choice between SEVI and LEVI

In empirical applications, it is often the case that neither the SEVI nor the LEVI model is correctly specified. However, we still have to decide which model to use. Since both models have

the same number of parameters, we can compare their empirical (quasi) negative log-likelihoods, which leads to the negative log-likelihood difference given by

$$\begin{aligned}\text{LR} &:= \ell_{\text{SEVI}}(\hat{\beta}_{\text{SEVI}}) - \ell_{\text{LEVI}}(\hat{\beta}_{\text{LEVI}}) \\ &= \sum_{i=1}^{n} \left[\ell_i^{\text{SEVI}}(\hat{\beta}_{\text{SEVI}}) - \ell_i^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})\right].\end{aligned}$$

Note that in Section 4, $\hat{\beta}_{\text{SEVI}}$ and $\hat{\beta}_{\text{LEVI}}$ were denoted as $\hat{\beta}_{\text{MLE}}$ and $\hat{\beta}_{\text{QMLE}}$, respectively. However, in this subsection, both the SEVI model and the LEVI model can be misspecified, and so both $\hat{\beta}_{\text{SEVI}}$ and $\hat{\beta}_{\text{LEVI}}$ can be quasi-maximum likelihood estimators.

The choice between the SEVI model and the LEVI model can then be based on LR. If $\text{LR} > 0$, then the LEVI model is preferred, because $\ell_{\text{SEVI}}(\cdot)$ and $\ell_{\text{LEVI}}(\cdot)$ are negative log-likelihoods, so a positive LR means that the LEVI model fits the data better. Conversely, if $\text{LR} < 0$, then the SEVI model is preferred. This selection rule is based on the empirical goodness of fit. With the parameter counts equal, it coincides with selection by the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC).

It is worth being explicit about what this comparison does and does not deliver. What it delivers is a ranking that is well defined whether or not either model is correctly specified, and one that involves no tuning choices, since the equal parameter counts leave nothing to trade off against fit. What it does not deliver is a test. The rule compares two point estimates of the empirical (quasi) log-likelihood, so it carries no size or power guarantee and it does not tell us whether the gap between the two fitted models exceeds what sampling variability alone would produce. A formal test is available: the two models are non-nested, and Vuong (1989) provides a likelihood-ratio test for exactly this setting. We deliberately report the simpler ranking, which requires no auxiliary variance estimate and no adjustment for the possibility that the two specifications are observationally equivalent. Accordingly, we read a negative LR as evidence that the SEVI specification is a serious contender for the data at hand, not as a rejection of the LEVI specification.

5 Simulation Evidence

We generate the data from a conditional SEVI model. More specifically, we assume that

$$Y_i = \arg\max_{j=1,\ldots,J} \{X_{ij}\beta_0 + \varepsilon_{ij}\},$$

where $\varepsilon_{ij} \sim iid$ SEVI over j. We set the number of covariates L in X_{ij} to be 3 and let $\beta_0 = (1, 2, 1)'$, and we consider two data-generating processes (DGPs) for the attributes $\{X_{ij}\}$. In the first DGP, for each i and j, we let $X_{ij,\ell}$ be iid $\mathcal{N}(0, \pi^2\omega_j^2/36)$ over $\ell = 1, 2, 3$ where

$$\omega_j = \frac{j - (J+1)/2}{\sqrt{(J^2-1)/12}},\ j = 1, \ldots, J. \tag{26}$$

Here, $(\omega_1, \ldots, \omega_J)$ is simply a standardized version of the sequence $(1, 2, \ldots, J)$. In this case, since ω_j^2 is symmetric about $(J+1)/2$ and increasing in $|j - (J+1)/2|$, each attribute $X_{ij,\ell}$ becomes more dispersed when the alternative label j is closer to 1 or J. As a result, the mean choice probabilities (i.e., $E\Pr(Y_i = j|V_i)$) are not all the same across the alternatives. For odd J, the middle alternative has $\omega_j = 0$; its attributes are then identically zero, so it enters as a reference

Table 1: MLE and QMLE of the model parameters for various J

	True	$J=2$		$J=5$		$J=8$		$J=11$	
	$(\boldsymbol{\beta_0})$	**MLE**	**QMLE**	**MLE**	**QMLE**	**MLE**	**QMLE**	**MLE**	**QMLE**
	$X_{ij,\ell} \sim iid\ \mathcal{N}(0, \pi^2\omega_j^2/36)$								
β_1	1.00	1.01	1.01	1.00	1.37	0.98	1.48	1.02	1.64
β_2	2.00	1.95	1.95	2.01	2.75	1.99	2.99	1.98	3.17
β_3	1.00	0.99	0.99	1.00	1.37	1.01	1.49	0.98	1.57
	$X_{ij,\ell} \sim iid\ \mathcal{N}(0, \pi^2/36)$								
β_1	1.00	1.01	1.01	1.01	1.35	0.99	1.51	1.00	1.64
β_2	2.00	2.03	2.03	2.05	2.80	1.98	3.09	1.98	3.37
β_3	1.00	1.03	1.03	1.03	1.38	1.03	1.56	1.00	1.64

alternative with zero systematic utility. In the second DGP, for each i and j, we let $X_{ij,\ell}$ be iid $\mathcal{N}(0, \pi^2/36)$ over $\ell = 1, 2, 3$. In this case, the attributes have the same distribution across the alternatives, and the mean choice probabilities are the same across all alternatives. Under these two DGPs for $\{X_{ij}\}$, the variances of the systematic utility $V_{ij}(\beta_0) = X_{ij}\beta_0$ are $\pi^2\omega_j^2/6$ and $\pi^2/6$, respectively. These variances are comparable to $\pi^2/6$, the variance of the random component.

Given a simple random sample $\{Y_i, X_{i1}, \ldots, X_{iJ}\}_{i=1}^n$, we use the MLE and QMLE in Section 4 to estimate β_0. To reiterate, the MLE is based on the correct specification that $\varepsilon_{ij} \sim iid$ SEVI, and the QMLE is based on the incorrect specification that $\varepsilon_{ij} \sim iid$ LEVI.

5.1 Simulation with a Large Sample Size

We first consider the case where the sample size is large and there is only one simulation replication. The purpose of this exercise is to examine the magnitude of the misspecification bias. Table 1 reports the estimates when the sample size n is $10,000$. When $J = 2$, the MLE and QMLE give the same estimate of β_0. This is consistent with the result that the SEVI and LEVI models are the same when there are only two alternatives. When $J > 2$, the MLE and QMLE are quite different. While the MLE is close to the true parameter vector for all values of J considered, the magnitude of the QMLE grows with J. For example, when $J = 8$ and $X_{ij,\ell} \sim iid\ \mathcal{N}(0, \pi^2\omega_j^2/36)$, the QMLE is $(1.48, 2.99, 1.49)$, each element of which is higher than the corresponding true value by about 50%. Thus, the misspecification bias of the QMLE is substantial. In line with Theorem 5, the ratio of the estimates of any two elements of β_0, say $\hat{\beta}_{\text{QMLE},2}/\hat{\beta}_{\text{QMLE},1}$, is close to the true ratio $\beta_{0,2}/\beta_{0,1}$. This is true regardless of whether the attributes have the same variance across the alternatives or not. Theorem 5 permits variance heterogeneity across the alternatives.

In a nonlinear model, instead of focusing on the parameter values, it can be more informative to examine the (empirical) average partial effects, defined by

$$APE(j;\beta,\text{SEVI}) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial P_{ij}^{\text{SEVI}}(\beta)}{\partial X_{ij,1}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial P_{ij}^{\text{SEVI}}(\beta)}{\partial v_{ij}(\beta)}\beta_1,$$

$$APE(j;\beta,\text{LEVI}) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial P_{ij}^{\text{LEVI}}(\beta)}{\partial X_{ij,1}} = \frac{1}{n}\sum_{i=1}^{n}P_{ij}^{\text{LEVI}}(\beta)\left(1 - P_{ij}^{\text{LEVI}}(\beta)\right)\beta_1.$$

In the above, P_{ij}^{SEVI} and P_{ij}^{LEVI} are the choice probabilities under the SEVI and LEVI specifications, respectively. $APE(j;\beta,\text{SEVI})$ and $APE(j;\beta,\text{LEVI})$ are the corresponding empirical

APEs that measure the effect on the probability of choosing alternative j when a marginal change is made to one of alternative j's attributes. We focus on the first attribute $X_{ij,1}$ and plot the ratios:

$$\frac{APE(j;\hat{\beta}_{\text{MLE}},\text{SEVI})}{APE(j;\beta_0,\text{SEVI})} \text{ and } \frac{APE(j;\hat{\beta}_{\text{QMLE}},\text{LEVI})}{APE(j;\beta_0,\text{SEVI})}$$

against the label of each alternative $j = 1,\ldots,J$. Figure 5 reports the results for the case that $X_{ij,\ell} \sim iid\ \mathcal{N}(0,\pi^2\omega_j^2/36)$ and makes it clear that the empirical APE is not close to the target APE (i.e., $APE(j;\beta_0,\text{SEVI})$) when the model is misspecified. This shows that the difference in the parameter estimates due to model misspecification has substantial effects on APE estimation.

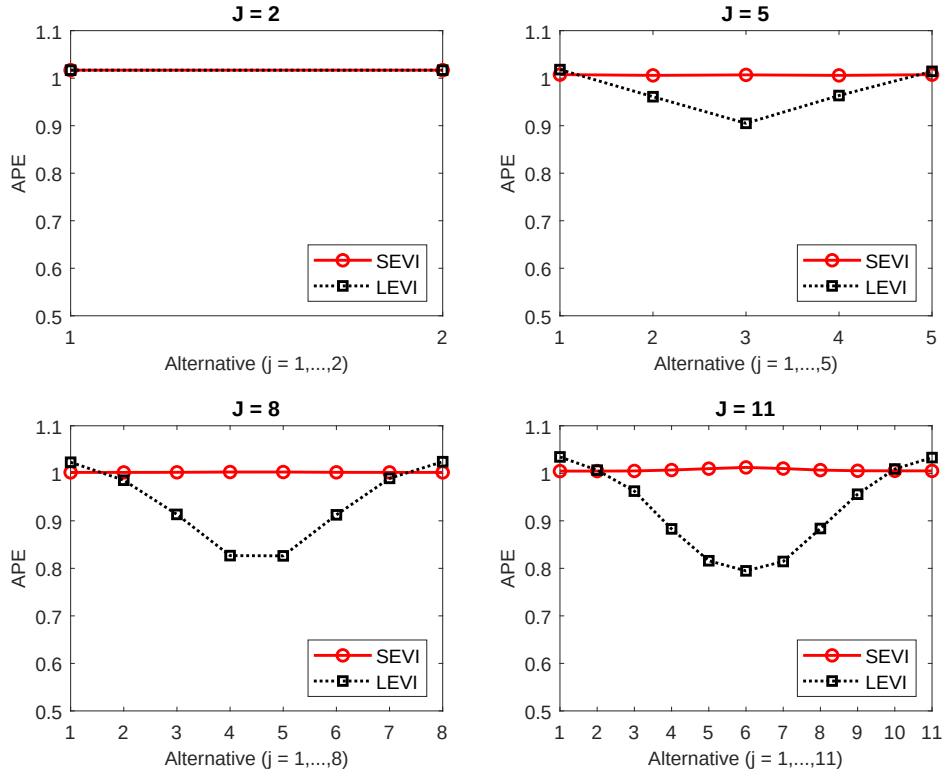


Figure 5: Plots of the empirical APE under the correct SEVI specification and the incorrect LEVI specification against alternative labels for a sample of $n = 10,000$ individuals

To further illustrate the difference between the SEVI and LEVI specifications, we conduct an out-of-sample prediction exercise. Let $\hat{\beta}^{(J)}_{\text{MLE}}$ and $\hat{\beta}^{(J)}_{\text{QMLE}}$ be the MLE and QMLE of β_0, respectively, when there are J alternatives. We use these estimates to compute out-of-sample choice probabilities for a certain choice problem. More specifically, we assume that an out-of-sample individual chooses between the first two alternatives only, while the remaining $J-2$ alternatives are unavailable. In this case, the out-of-sample choice probabilities are

$$\hat{P}_{i_o,\text{MLE}} := P\left(Y_{i_o} = 1|X_{i_o};\hat{\beta}^{(J)}_{\text{MLE}}\right) \text{ and } \hat{P}_{i_o,\text{QMLE}} := P\left(Y_{i_o} = 1|X_{i_o};\hat{\beta}^{(J)}_{\text{QMLE}}\right)$$

where

$$P\left(Y_{i_o} = 1|X_{i_o};\beta\right) = \frac{\exp(X_{i_o,1}\beta)}{\exp(X_{i_o,1}\beta)+\exp(X_{i_o,2}\beta)}$$

and $i_o \in \{n+1, \ldots, 2n\}$ represents an out-of-sample individual whose observation was not used in constructing the MLE or QMLE, while X_{i_o} is generated in the same way as any in-sample observation. Note that the number of "out-of-sample" individuals is the same as the number of "in-sample" individuals, and both are equal to 10,000.

When there are only two alternatives in the out-of-sample prediction exercise, the formulae for the choice probabilities are the same under SEVI or LEVI. Hence, as given above, $\hat{P}_{i_o,\text{MLE}}$ and $\hat{P}_{i_o,\text{QMLE}}$ are different only because they are based on different estimated parameters.

Figure 6 reports the kernel density estimators (KDEs) of the out-of-sample predictive probabilities for $\beta = \beta_0, \hat{\beta}^{(J)}_{\text{MLE}}$ and $\hat{\beta}^{(J)}_{\text{QMLE}}$ when $X_{ij,\ell} \sim iid\, \mathcal{N}(0, \pi^2\omega_j^2/36)$. Panels (a) and (c) plot the KDEs of $\{P(Y_{i_o} = 1|X_{i_o}; \beta)\}_{i_o=n+1}^{2n}$, and panels (b) and (d) those of $\{P(Y_{i_o} = 2|X_{i_o}; \beta)\}_{i_o=n+1}^{2n}$; panels (a) and (b) take $J = 5$ and panels (c) and (d) take $J = 11$. Unsurprisingly, the KDE under $\hat{\beta}^{(J)}_{\text{MLE}}$ nearly coincides with the target KDE, the one under β_0: at this sample size $\hat{\beta}^{(J)}_{\text{MLE}}$ is close to β_0, as Table 1 demonstrates, so the two sets of predicted probabilities, and hence their densities, nearly agree. On the other hand, the KDE under $\hat{\beta}^{(J)}_{\text{QMLE}}$ is quite different from the target KDE. Thus, the misspecification biases the parameter estimates and, in turn, distorts the predicted choice probabilities.

The corresponding figure for $X_{ij,\ell} \sim iid\, \mathcal{N}(0, \pi^2/36)$ is qualitatively similar and is thus omitted. The difference between the KDE under $\hat{\beta}^{(J)}_{\text{QMLE}}$ and the target KDE appears to increase with J.

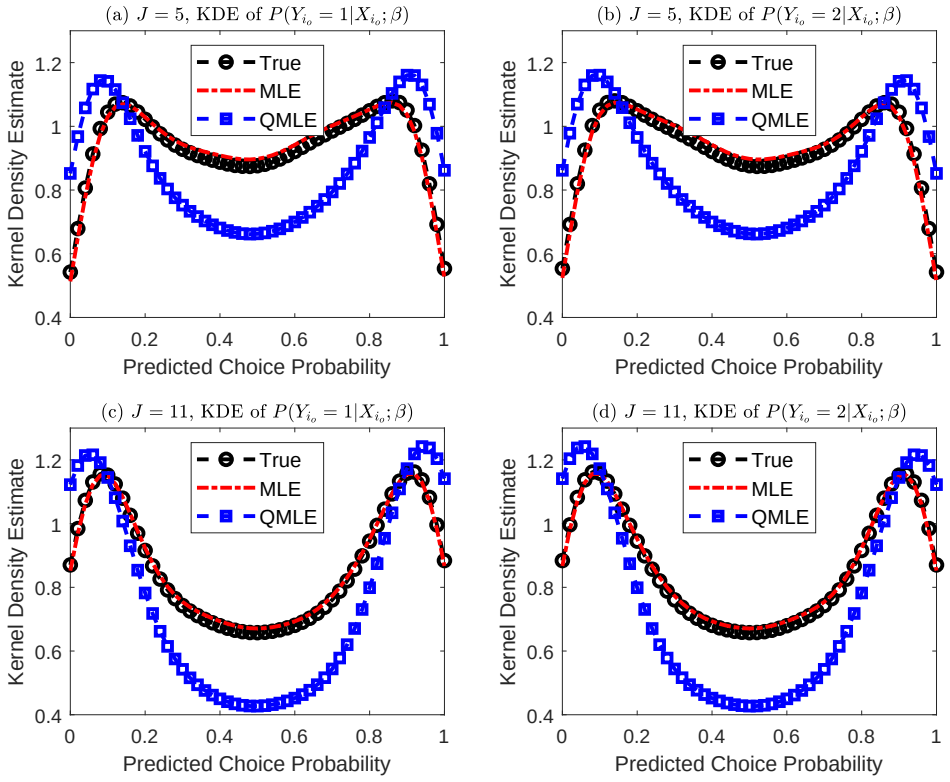


Figure 6: Kernel density estimation of the out-of-sample predictive probabilities when $n = 10,000, J = 5, 11$ and $X_{ij,\ell} \sim iid\mathcal{N}(0, \pi^2\omega_j^2/36)$

5.2 Simulation with Smaller Sample Sizes

In this subsection, we study the performance of the MLE and QMLE in finite samples. Using the same DGP as in the previous subsection, we consider the sample sizes $n = 500$ and $1,000$, and the number of alternatives $J = 2, 5, 8, 11$. The number of simulation replications is 5,000. For each estimator, we compute the bias, the empirical standard deviation, and the average of the standard errors across the simulation replications. Furthermore, we construct a 95% confidence interval (CI) for each element of β_0 following the standard procedure of adding and subtracting 1.96 times the standard error, and we report the empirical coverage of the CI.

Table 2 reports the results when $X_{ij,\ell} \sim iid\ \mathcal{N}(0, \pi^2\omega_j^2/36)$ and $n = 500$. The reported results are representative of other cases with a different DGP for $X_{ij,\ell}$ or sample size. A few patterns emerge. First, the bias of the MLE is small for every J, and what variation there is across J lies within simulation error. On the other hand, the bias of the QMLE is large and increases with J. This is consistent with our large sample result when $n = 10,000$ and there is only one simulation replication. Here, the finite sample biases (with $n = 500$) are averaged over 5,000 simulation replications. Second, for both the MLE and QMLE, the empirical standard deviation decreases with J. A larger J can be regarded as having more observations and, hence, lower sampling noise. Third, for both estimators, the average of the standard errors across the simulation replications matches the corresponding standard deviation very well. This indicates that the standard errors are reliable estimates of the standard deviation. Finally, for the MLE, the empirical coverage of the 95% confidence intervals is very close to 95%. This shows that the asymptotic normal approximation is reliable. On the other hand, for the QMLE, the empirical coverage of the 95% confidence intervals is much smaller than 95%. The only exception is the case when $J = 2$, where the MLE and QMLE are identical. In this case, there is a nearly invisible difference in the empirical coverages, because the standard error for the MLE is based on the non-robust asymptotic variance estimator, while the standard error for the QMLE is based on the robust sandwich asymptotic variance estimator. Both variance estimators are consistent when $J = 2$. When $J > 2$, the under-coverage of the CIs based on the QMLE is due to the bias in the QMLE. The larger the number of alternatives J is, the larger the bias becomes, and the smaller the empirical coverage is.

Next, we consider testing IIA using the case when $X_{ij,\ell} \sim iid\ \mathcal{N}(0, \pi^2/36), \ell = 1, 2, 3$ and $n = 500$ as an example. We follow the standard procedure to perform the Hausman–McFadden test (cf. Hausman and McFadden (1984)). For each simulation replication, we first estimate the model under the (incorrect) LEVI specification using the full sample. We then carry out the same estimation procedure but use only the subsample of individuals who choose either alternative 1 or 2. We compare these two estimates using the Wald statistic based on the 95% critical value from the chi-squared distribution with three degrees of freedom ($\chi^2(3)$). The empirical rates of rejection over 5,000 simulation replications for $J = 5$, 8, and 11 are 30.30%, 40.18%, and 47.52%, respectively. Since IIA genuinely fails under the SEVI design, these are rejection frequencies under the alternative, so the test misses the violation in more than half the replications even at $J = 11$. When it does fail, the standard practice is to employ discrete choice models that do not exhibit the IIA property, such as the mixed logit. Such models relax the specification of the systematic utility, allowing the preference parameters to vary across individuals without changing the distribution of the random component. As a result, the mixed logit does not address a misspecification in the skewness of the random component, which is the source of the rejection here.

Finally, we consider using the AIC/BIC to decide which model fits the data better. Since the

Table 2: Finite sample performance of the MLE and QMLE

	MLE				QMLE			
	Empirical Bias	Empirical StD	Average of SE's	Empirical Coverage	Empirical Bias	Empirical StD	Average of SE's	Empirical Coverage
				$J = 2$				
β_1	0.009	0.211	0.210	0.951	0.009	0.211	0.209	0.948
β_2	0.023	0.240	0.241	0.950	0.023	0.240	0.239	0.950
β_3	0.010	0.209	0.210	0.950	0.010	0.209	0.209	0.950
				$J = 5$				
β_1	0.003	0.105	0.107	0.957	0.366	0.141	0.142	0.261
β_2	0.010	0.136	0.137	0.950	0.738	0.175	0.173	0.004
β_3	0.004	0.106	0.107	0.958	0.367	0.142	0.142	0.266
				$J = 8$				
β_1	0.006	0.087	0.088	0.952	0.514	0.130	0.129	0.017
β_2	0.011	0.117	0.116	0.951	1.026	0.159	0.158	0.000
β_3	0.006	0.088	0.088	0.951	0.513	0.129	0.129	0.017
				$J = 11$				
β_1	0.004	0.080	0.080	0.950	0.600	0.124	0.123	0.000
β_2	0.007	0.105	0.107	0.955	1.201	0.149	0.150	0.000
β_3	0.007	0.081	0.080	0.946	0.605	0.126	0.123	0.001

two models have the same number of parameters, selection by the AIC or the BIC coincides with the rule based on the sign of the negative log-likelihood difference LR in Section 4.3. When the SEVI model is correctly specified, $X_{ij,\ell} \sim iid\ \mathcal{N}(0, \pi^2\omega_j^2/36)$ and $n = 500$, the fractions of the 5,000 replications in which LR < 0, so that the SEVI model is selected, are 92.56%, 97.86%, and 99.32% for $J = 5$, 8, and 11, respectively. Results are similar when $X_{ij,\ell} \sim iid\ \mathcal{N}(0, \pi^2/36)$ and for other sample sizes. On the other hand, when the LEVI model is correctly specified, $X_{ij,\ell} \sim iid$ $\mathcal{N}(0, \pi^2\omega_j^2/36)$ and $n = 500$, the fractions in which LR > 0, so that the LEVI model is selected, are 87.56%, 94.94%, and 97.18% for $J = 5$, 8, and 11, respectively. Therefore, regardless of which of the two models generated the data, the AIC and BIC perform very well in identifying it when one of them is correctly specified.

5.3 Computation Time Comparison

We compare the times required to estimate the LEVI, SEVI, and NORM ($\mathcal{N}(0, \pi^2/6)$) models on data generated with SEVI errors, fitting all three with a correctly specified systematic utility. The SEVI choice probabilities are computed from the formula in Lemma 1, with the choice subsets enumerated by Gosper's hack. The attributes are drawn as in the first DGP above, but the designs depart from it in three ways: (i) J runs from 2 to 16 at $n = 1,000$, (ii) designs with ten attributes are run alongside those with three, and (iii) β_0 is drawn from a multivariate standard normal distribution rather than fixed. A second group of designs adds alternative-specific constants, either alone or alongside three attributes. For $J \leq 12$, the SEVI model takes about as much time as the LEVI model: under 27 seconds with ten attributes at $J = 12$, and under 65 seconds at the same J in the designs with alternative-specific constants. Over the same range, the NORM model is far slower. At $J = 10$ and $J = 12$, the required times for the NORM

model are 2,274 and 5,159 seconds—between half an hour and two hours—in the design whose only covariates are alternative-specific constants, whereas the SEVI model takes about a minute. Beyond $J = 12$, the computation time for the SEVI model rises steeply and approaches that of the NORM model by $J = 16$. Section S.H.3 of the supplementary appendix gives the designs, the implementation, and the two figures that plot the computation time against the number of alternatives.

6 Empirical Applications

In this section, we examine four commonly used datasets (with the last one containing three sub-datasets) from well-cited textbooks and papers, all of which are included in the mlogit package in R (cf. Croissant (2020)). For each dataset, we consider three different specifications of the random components: LEVI, SEVI, and NORM (i.e., $\mathcal{N}\left(0, \pi^2/6\right)$). The variances of the random components for all three models are identical and equal to $\pi^2/6$. Since the means of the random components do not matter, the main difference among these three specifications lies in the shape of the distributions, starting with their skewness.[11]

The four datasets we analyze are:

1. Fishing mode choice: the dataset comes from Herriges and Kling (1999)and has been analyzed in several textbooks, including Cameron and Trivedi (2005), Cameron (2022), and Cameron and Trivedi (2022).

2. Stated preference vehicle choice: the dataset originates from Brownstone et al. (1996) and is subsequently used in Brownstone and Train (1998) and McFadden and Train (2000).

3. Choice of brand for saltine crackers: the dataset comes from Paap and Franses (2000), and a subset of the data were analyzed in Jain et al. (1994).

4. Choice of NO_x emissions compliance strategies: the dataset originates from Fowlie (2010).

The last application in our analysis, the NO_x emissions study by Fowlie (2010), differs fundamentally from the first three. While the first three applications follow a standard utility-maximization framework, the last one centers on cost minimization. To model this behavior, we assume that there is a latent function C_{ij} that encompasses an observable component $D_{ij}\left(\beta_0\right)$ and an unobservable component ε_{ij}, so that

$$C_{ij}\left(\beta_0\right) = D_{ij}\left(\beta_0\right) + \varepsilon_{ij} \text{ for } j = 1, \ldots, J \tag{27}$$

where $C_{ij}\left(\beta_0\right)$ is the cost that individual i incurs from choosing alternative j. The observed choice of individual i, denoted by Y_i, is

$$Y_i = \arg \min_{j=1,\ldots,J} C_{ij}\left(\beta_0\right). \tag{28}$$

The study by Anderson and de Palma (1999) appears to be motivated by the above cost-minimization problem. It is easy to show that

$$\begin{aligned}
&\Pr(Y_i = j | D_i\left(\beta_0\right) = d_i, \varepsilon_{im} \overset{iid}{\sim} \text{LEVI}) = P^{\text{SEVI}}\left(Y_i = j | -d_i\right), \\
&\Pr(Y_i = j | D_i\left(\beta_0\right) = d_i, \varepsilon_{im} \overset{iid}{\sim} \text{SEVI}) = P^{\text{LEVI}}\left(Y_i = j | -d_i\right),
\end{aligned}$$

[11] Note that we are conducting a comparative analysis solely among the three specifications, and we do not claim that any of these specifications is consistent with the true underlying DGP.

Table 3: Performance of LEVI, SEVI, and NORM on several datasets

Dataset	n	J	L	Computation time			Log-likelihood			Likelihood
				LEVI	SEVI	NORM	LEVI	SEVI	NORM	Ranking
Fishing	1,182	4	8	0.35	0.58	370	-1215.14	-1213.21	-1218.93	S > L > N
Vehicles	4,654	6	21	3.57	14.00	2,232	-7394.62	-7388.75	-7389.50	S > N > L
Crackers	3,289	4	6	0.31	0.55	157	-3347.61	-3347.13	-3344.51	N > S > L
NO_x: dereg	227	15	6	135.00	0.02	791	-345.35	-339.07	-343.21	S > N > L
NO_x: public	113	15	6	78.00	0.01	2,357	-86.30	-78.46	-82.38	S > N > L
NO_x: reg	292	15	6	279.00	0.01	4,600	-364.99	-359.74	-365.96	S > L > N

(i) n is the # of choice situations, J is the (max) # of alternatives, and L is the total # of covariates.
(ii) "NO_x: dereg," "NO_x: public," and "NO_x: reg" are the subsamples of deregulated, public, and regulated plants, respectively.
(iii) All computation times are measured in seconds.
(iv) In the likelihood ranking, S, L, and N are short for SEVI, LEVI, and NORM, respectively.

where $P^{\text{SEVI}}(Y_i = j| - d_i)$ and $P^{\text{LEVI}}(Y_i = j| - d_i)$ are defined in (3) and (5), respectively. So, for discrete choice problems aiming to minimize an objective function, when $\varepsilon_{ij} \overset{iid}{\sim}$ LEVI, the choice probability takes the form of (3). In contrast, when $\varepsilon_{ij} \overset{iid}{\sim}$ SEVI, the choice probability takes the usual logit form.

Note that there is a sign change: it is $-d_i$ rather than d_i itself that enters the expression of the choice probability derived under utility maximization. If we parametrize $D_{ij}(\beta_0) = X_{ij}\beta_0$ for some covariate X_{ij} and run a conditional logit using X_{ij} as the covariate in a standard statistical package, then we are assuming that $\varepsilon_{ij} \overset{iid}{\sim}$ SEVI, and what the package returns is an estimate of $-\beta_0$. To correct this, we can flip the sign of the estimate or use the negated covariate $-X_{ij}$.

If we deem that $\varepsilon_{ij} \overset{iid}{\sim}$ LEVI is more plausible, then the choice probability formula in (3) has to be used, and the approach proposed in this paper can be easily applied.

Irrespective of whether we are dealing with a minimization or maximization choice problem, in discussing the empirical applications we refer to the model with SEVI errors and the model with LEVI errors, naming in each case the distribution of the random component as it enters the objective function as written. In the case of maximization problems, the model with LEVI errors is more straightforward to estimate, whereas for minimization problems, the model with SEVI errors is easier to estimate using standard software.

Table 3 presents the computation times and log-likelihood performance of the three models. In all cases, the computation time represents the seconds spent by our MATLAB program to find the MLE/QMLE. Although there may be slight fluctuations in the reported computation times due to concurrent background processes, these times serve as a reliable indicator of the time needed to estimate each model. We will revisit this table when we discuss each application.

6.1 Application: Choice of Fishing Mode

For the fishing mode application, we model an individual's choice of fishing modes among four mutually exclusive alternatives: beach ($j = 1$), pier ($j = 2$), private boat ($j = 3$), and charter

boat ($j = 4$). We assume that the latent utility is given by

$$U_{ij} = X_{1ij}\beta_{x1} + X_{2ij}\beta_{x2} + D2_j \times \beta_{d2} + D3_j \times \beta_{d3} + D4_j \times \beta_{d4} + (Z_i \cdot D2_j) \times \beta_{z2} + (Z_i \cdot D3_j) \times \beta_{z3} + (Z_i \cdot D4_j) \times \beta_{z4} + \varepsilon_{ij},$$

where X_{1ij} and X_{2ij} are the price and catch rate of alternative j faced by individual i, $D2_j, D3_j$, and $D4_j$ are the dummy variables for alternatives 2, 3, and 4, respectively, and Z_i is the income of individual i. For this application, the models with LEVI and SEVI errors yield similar parameter estimates and likelihood function values. The correlation coefficient between the in-sample predicted choice probabilities for the two models is 0.998, indicating a nearly perfect positive correlation between them. Even so, the model with SEVI errors has a higher log-likelihood than the model with LEVI errors. Both models can be estimated with minimal computational cost, with the optimization algorithm completing in less than a minute. In contrast, the multinomial probit incurs significantly higher computational costs and yields a lower log-likelihood than both the SEVI and LEVI models.

Figure S.10 in the supplementary appendix plots the difference between the predicted choice probabilities (i.e., $P^{\text{SEVI}}(\hat{\beta}_{\text{SEVI}}) - P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$ or $P^{\text{NORM}}(\hat{\beta}_{\text{NORM}}) - P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$) against $P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$. While all three predicted choice probabilities are very highly correlated, the figure illustrates some subtle distinctions. Specifically, as $P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$ increases, the variabilities of $P^{\text{SEVI}}(\hat{\beta}_{\text{SEVI}}) - P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$ and $P^{\text{NORM}}(\hat{\beta}_{\text{NORM}}) - P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$ also increase, indicating that the three models tend to align closely when the choice probabilities are low but exhibit discrepancies otherwise.

To illustrate the difference between the models with SEVI and LEVI errors using the fishing choice data, we calculate the choice probabilities $P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$ and $P^{\text{SEVI}}(\hat{\beta}_{\text{LEVI}})$ using the same parameter value $\hat{\beta}_{\text{LEVI}}$. Figure S.11 in the supplementary appendix plots $P^{\text{SEVI}}(\hat{\beta}_{\text{LEVI}})$ against $P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$. We find that when $P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$ is small, $P^{\text{SEVI}}(\hat{\beta}_{\text{LEVI}})$ tends to be smaller than $P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$, and when $P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$ is large, $P^{\text{SEVI}}(\hat{\beta}_{\text{LEVI}})$ tends to be larger than $P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$. This observation aligns with the simulation results reported in Section S.H of the supplementary appendix, as seen in Figure S.6, for example.

6.2 Application: Stated Preferences for Vehicle Choice

For the stated-preference vehicle data, each individual was presented with 6 hypothetical vehicles fueled by methanol, compressed natural gas, electricity, or regular gas. These vehicles have various attributes, such as fuel type, price, cost per mile, mileage range between refuelings/rechargings, vehicle type (SUV, sports car, station wagon, truck, van, regular car), pollution level, and speed, among others. We employ the same set of 21 covariates as in McFadden and Train (2000) (Table 1), some of which are the products of vehicle attributes and individual characteristics, such as household size greater than 2 or not, college education or not, and a daily commute of less than 5 miles or not.

As shown in Table 3, in terms of the log-likelihood achieved, the model with LEVI errors performs worse than the model with SEVI errors and the NORM model.

Figures 7 (a) and (b) plot the difference in the predicted probabilities (i.e., $P^{\text{SEVI}}(\hat{\beta}_{\text{SEVI}}) - P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$ or $P^{\text{NORM}}(\hat{\beta}_{\text{NORM}}) - P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$) against the LEVI predicted probability (i.e., $P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$). There are some intriguing patterns. Figure 7(a) shows that $P^{\text{SEVI}}(\hat{\beta}_{\text{SEVI}})$ tends to be higher than $P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$ when the latter is low and tends to be lower than $P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$ when the latter is high.

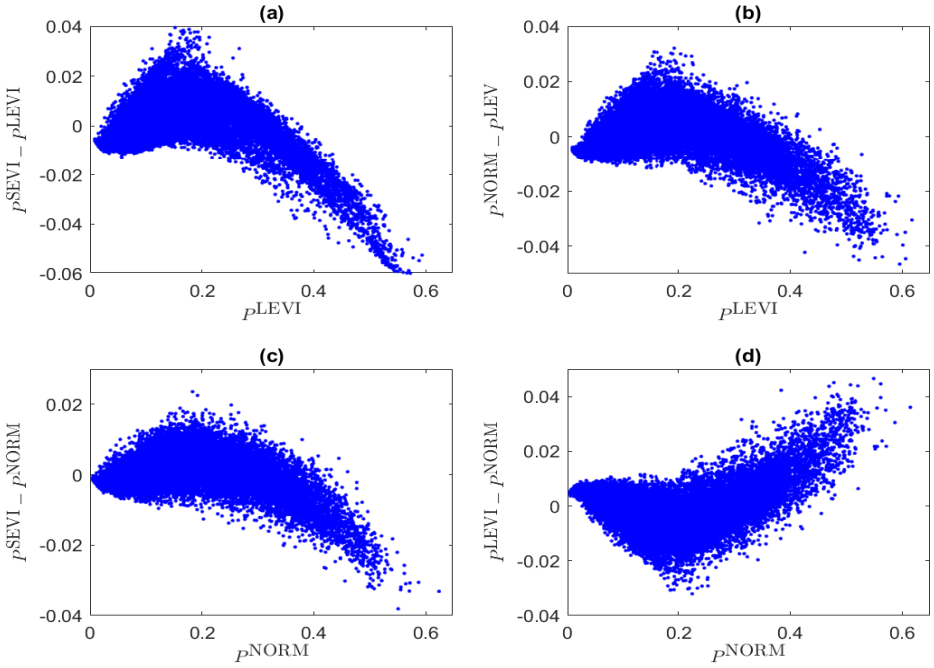


Figure 7: Plot of the difference of the predicted choice probabilities against the LEVI or NORM predicted probability in the vehicle choice application

Note that the difference $P^{\text{SEVI}}(\hat{\beta}_{\text{SEVI}}) - P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$ can be decomposed as

$$P^{\text{SEVI}}(\hat{\beta}_{\text{SEVI}}) - P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}}) = \underbrace{P^{\text{SEVI}}(\hat{\beta}_{\text{SEVI}}) - P^{\text{SEVI}}(\hat{\beta}_{\text{LEVI}})}_{\text{I}} + \underbrace{P^{\text{SEVI}}(\hat{\beta}_{\text{LEVI}}) - P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})}_{\text{II}}.$$

The first part (I) above is due to the difference in the parameter estimates used ($\hat{\beta}_{\text{SEVI}}$ vs. $\hat{\beta}_{\text{LEVI}}$), and the second part (II) above is due to the difference in the predictive models used (SEVI vs. LEVI). According to an unreported figure, the second part exhibits the same pattern as in Figure S.6. So the pattern in Figure 7(a) emerges because the effect of the first part dominates that of the second part. Figure 7(b) shows a similar pattern, although $P^{\text{NORM}}(\hat{\beta}_{\text{NORM}}) - P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$ appears to be less variable than $P^{\text{SEVI}}(\hat{\beta}_{\text{SEVI}}) - P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$.

Figures 7 (c) and (d) are similar to Figures 7 (a) and (b) but use $P^{\text{NORM}}(\hat{\beta}_{\text{NORM}})$ as the benchmark. According to these two figures, $P^{\text{SEVI}}(\hat{\beta}_{\text{SEVI}}) - P^{\text{NORM}}(\hat{\beta}_{\text{NORM}})$ and $P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}}) - P^{\text{NORM}}(\hat{\beta}_{\text{NORM}})$ display opposite patterns. On the one hand, $P^{\text{SEVI}}(\hat{\beta}_{\text{SEVI}})$ tends to be higher than $P^{\text{NORM}}(\hat{\beta}_{\text{NORM}})$ when the latter is low and tends to be lower than $P^{\text{NORM}}(\hat{\beta}_{\text{NORM}})$ when the latter is high. On the other hand, $P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$ tends to be lower than $P^{\text{NORM}}(\hat{\beta}_{\text{NORM}})$ when the latter is low and tends to be higher than $P^{\text{NORM}}(\hat{\beta}_{\text{NORM}})$ when the latter is high. The choice among SEVI, NORM, and LEVI random components can therefore have a substantial effect on the predicted choice probabilities.

We also report the point estimates of the parameters under the LEVI and SEVI models, along with their 95% confidence intervals. See Figure S.12 in the supplementary appendix. While the CIs for 14 parameters clearly overlap, those for other parameters either do not overlap or barely overlap. In practical terms, this divergence suggests that the two models may lead to different

conclusions regarding the impact of certain attributes on vehicle choices. Specifically, attributes such as mileage range between refuelings/rechargings, cost per mile, and vehicle types exhibit different effects in the models with LEVI and SEVI errors. Compared to the model with LEVI errors, the model with SEVI errors suggests that the mileage range between refuelings/rechargings and the vehicle's classification as an SUV (rather than a regular car) are less influential factors. On the other hand, the cost per mile and other vehicle types (rather than a regular car) appear to be more influential in the model with SEVI errors.

6.3 Application: Choice of Saltine Crackers

The dataset on saltine cracker choices contains information on all 3,289 cracker purchases by 136 households, including the brand chosen from four alternative brands, prices of all four brands (`Price`), whether there was an in-store display of the brand (`Display` dummy) and/or whether the brand was featured in the store's newspaper advertising (`Feature` dummy) at the time of purchase. The latent utility is specified as follows:

$$\begin{aligned} U_{ij} = {} & \texttt{Price}_{ij} \times \beta_{x1} + \texttt{Display}_{ij} \times \beta_{x2} + \texttt{Feature}_{ij} \times \beta_{x3} \\ & + D2_j \times \beta_{d2} + D3_j \times \beta_{d3} + D4_j \times \beta_{d4} + \varepsilon_{ij}, \end{aligned}$$

where $j = 1, 2, 3, 4$ correspond to the brands "Sunshine," "Keebler," "Nabisco," and "Private" (a private-label brand), respectively, and $D2$, $D3$, and $D4$ are dummies for the latter three brands (the dummy for the first brand has been omitted to achieve identification).

For this application, individuals make repeated purchase decisions. Choice situations belonging to the same individual may be correlated. To account for the correlation, we treat each individual as a cluster and employ cluster-robust standard errors. An unreported figure similar to Figure S.12 shows that the estimated coefficients on the three main covariates (`Price`, `Display`, and `Feature`) are similar across the models with LEVI and SEVI errors. However, the alternative-specific coefficients on $D2$, $D3$, and $D4$ are different across these two models. As shown in Table 3, the model with SEVI errors has a slightly higher log-likelihood than the model with LEVI errors, although the difference is small.

Table 3 also shows that the NORM model attains a higher log-likelihood than both the model with LEVI errors and the model with SEVI errors. This may indicate that the random utility component is symmetrically distributed, so that a normal distribution might provide a more accurate characterization than the LEVI and SEVI distributions. Figure S.13 in the supplementary appendix plots the difference in the predicted choice probabilities using the NORM probability as the benchmark, highlighting distinct deviations of the SEVI and LEVI probabilities from the NORM probability.

6.4 Application: Choice of a NO_x Emissions Reduction Technology

Managers of coal-fired electric power plants choose a NO_x emissions reduction technology in order to comply with a regional NO_x emissions trading program. The manager of a plant unit, say unit i, chooses a compliance strategy from J_i alternatives indexed by $j = 1, \ldots, J_i$ to minimize the unobserved overall cost C_{ij}. We use the model in (27) and (28), but there are two differences. First, the maximum number of options is 15, but not all options are available for each plant. To accommodate this, we simply assign an infinite systematic cost to any unavailable option, ensuring it is never selected. This is not computationally efficient, but it makes the number of

options the same across all choice situations. In addition, this method effectively provides us with an empirical example in which the number of alternatives is quite large.

Padding the choice set in this way is exact rather than approximate. Let $\mathcal{A}_i \subseteq \mathcal{J}$ collect the options that are actually available to unit i, and take $j \in \mathcal{A}_i$. In the equivalent maximization representation, an infinite systematic cost for an unavailable option $k \notin \mathcal{A}_i$ corresponds to a systematic utility of $-\infty$, so that its scale value $s_{ik} = \exp(v_{ik})$ tends to zero and $s^i_{j|k} = s_{ij}/s_{ik} \to \infty$. Every term of the alternating sum (4) whose index set $\mathcal{S}$ contains such a k therefore satisfies

$$\frac{1}{1 + \sum_{r \in \mathcal{S}} s^i_{j|r}} \longrightarrow 0,$$

and the sum collapses onto the subsets of $\mathcal{A}_i \setminus \{j\}$. In the limit, the padded choice set therefore delivers the same choice probability, and hence the same likelihood contribution, as the unpadded one. The argument applies verbatim to whichever of the two models carries the extended choice probability formula.

Second, some plants have the same manager, so there may be dependence in the unobserved costs across these plants. Following standard practice, we do not model the dependence when constructing the likelihood function. That is, we use a partial-likelihood approach, but we account for the dependence in estimating the asymptotic variance by clustering the managers.

The compliance options have different capital costs (denoted by $\texttt{kcost}_{ij}$) and variable operating costs (denoted by $\texttt{vcost}_{ij}$). Each compliance option incorporates one or more of the following three technologies: post-combustion pollution control technology (indicated by the dummy $\texttt{post}_j$), combustion modification technology (indicated by the dummy $\texttt{cm}_j$), and low NO_x burners technology (indicated by the dummy $\texttt{lnb}_j$). Taking these into consideration, we follow Fowlie (2010) and let $X_{ij} = (\texttt{post}_j, \texttt{cm}_j, \texttt{lnb}_j, \texttt{kcost}_{ij}, \texttt{vcost}_{ij}, \texttt{kage}_{ij})$ where $\texttt{kage} = \texttt{kcost} \times \texttt{age}$ and $\texttt{age}$ is the age of a plant.

Each plant is in one of three regulatory environments: "deregulated," "public," or "regulated." The central question investigated by Fowlie (2010) is whether and how regulatory environments influence a manager's choice of compliance strategies. The question can be addressed by comparing the coefficient estimates across the subsamples of "deregulated," "public," and "regulated" plants.

We apply the models with LEVI, SEVI, and normal errors to each subsample and evaluate their performance. Table 3 shows that the model with SEVI errors, which assumes that $\varepsilon_{ij} \overset{iid}{\sim}$ SEVI, achieves the highest likelihood among the three models for all three subsamples. We note that Fowlie (2010) made this SEVI assumption implicitly and, in doing so, happened to use the model with the highest likelihood among the three.

The parameter estimates and the cluster-robust standard errors are reported in Table 4. For each of the three models, the coefficients vary significantly across different regulatory environments. In particular, the coefficient on capital cost ($\texttt{kcost}$) based on the "deregulated" subsample is positive and statistically significant at the 1% level, but those based on the "public" subsample and the "regulated" subsample are negative, albeit not statistically significant. This implies that plant managers give more weight to capital cost if their plants are in a "deregulated" market rather than a "public" or "regulated" market. Our findings confirm the main result of Fowlie (2010) and show that the result is robust to specifying the random cost component as SEVI, LEVI, or normal.

Note that for the same subsample, the coefficients in the models with LEVI and SEVI errors may be directly compared, since the two error distributions share the variance $\pi^2/6$. For the

Table 4: Parameter estimates and standard errors: environmental compliance choices

	$\hat{\beta}_{\text{LEVI}}$	se:$\hat{\beta}_{\text{LEVI}}$	$\hat{\beta}_{\text{SEVI}}$	se:$\hat{\beta}_{\text{SEVI}}$	$\hat{\beta}_{\text{NORM}}$	se:$\hat{\beta}_{\text{NORM}}$
			Deregulated			
`post`	1.502	0.106	0.862	0.056	1.245	0.177
`cm`	1.538	0.120	0.859	0.068	1.250	0.155
`lnb`	1.551	0.121	0.784	0.064	1.193	0.175
`vcost`	0.188	0.045	0.112	0.030	0.145	0.041
`kcost`	0.060	0.020	0.036	0.012	0.050	0.018
`kage`	0.037	0.008	0.028	0.004	0.037	0.010
			Public			
`post`	5.706	1.050	3.890	0.601	4.927	0.824
`cm`	4.433	0.442	2.685	0.193	3.589	0.384
`lnb`	3.964	0.307	2.532	0.201	3.323	0.463
`vcost`	1.564	0.604	0.840	0.293	1.177	0.270
`kcost`	-0.039	0.255	-0.100	0.118	-0.086	0.088
`kage`	0.080	0.046	0.024	0.027	0.049	0.038
			Regulated			
`post`	2.665	0.142	1.680	0.070	2.302	0.213
`cm`	1.911	0.103	1.250	0.058	1.686	0.151
`lnb`	2.208	0.111	1.377	0.057	1.884	0.180
`vcost`	0.278	0.045	0.171	0.028	0.196	0.047
`kcost`	-0.008	0.023	-0.005	0.013	-0.023	0.022
`kage`	0.023	0.005	0.014	0.003	0.010	0.008

Note: The reported coefficients are those of the systematic cost component $D_{ij}(\beta_0)$ in (27). A positive coefficient means that the attribute raises the overall cost of an alternative and therefore makes that alternative less likely to be chosen.

"deregulated" subsample, each element of $\hat{\beta}_{\text{LEVI}}$ is larger than the corresponding element of $\hat{\beta}_{\text{SEVI}}$ by more than 30%. For the "public" and "regulated" subsamples, the coefficients on `post`, `cm`, `lnb`, `vcost`, and `kage` in the model with LEVI errors are larger than those in the model with SEVI errors by at least 45%. In the "regulated" subsample the inflation factors are quite similar, ranging only from 1.53 to 1.64, which is what Theorem 5 leads one to expect: under misspecification the QMLE converges to β_0 up to a single multiplicative factor, so ratios are far better preserved than levels. The factors vary more widely in the other two subsamples, which is unsurprising given that Assumption (iii) of the theorem is unlikely to hold for the binary technology indicators. This shows that the choice of model materially changes the parameter estimates, with implications for policy recommendations.

6.5 Further Remarks on the Empirical Applications

We conclude the empirical applications by making some further comments on the rankings of the three models according to their likelihood values and computation times.

First, in terms of likelihood values, the model with SEVI errors attains a higher likelihood than the model with LEVI errors across all the applications considered. With the exception of the cracker brand application, the model with SEVI errors achieves the highest likelihood value among all three models. Notably, in the environmental-compliance application, the model with SEVI errors yields a choice probability in the conventional logit form, because the agent minimizes rather than maximizes the objective. Overall, the model with SEVI errors attains a higher likelihood than the model with LEVI errors in both the utility-maximization and the

cost-minimization applications.

Second, for all applications, the computation time is shortest for the model whose choice probability formula takes the standard (closed-form) logit form: the model with LEVI errors in the utility-maximization applications, and the model with SEVI errors in the cost-minimization (environmental-compliance) application. When the number of alternatives, J, is not larger than 6, the computation time for the model with the extended choice probability formula, as given in (4), is less than four times as large as that for the computationally simplest model and is much lower than that for the multinomial probit model. When $J = 15$, the computation time for the model with the extended choice probability formula remains substantially lower than that of the multinomial probit model, although the margin narrows relative to smaller J.

Third, in every application the parameter estimates obtained under the model whose choice probability takes the conventional logit form—the model with LEVI errors in the utility-maximization applications and the model with SEVI errors in the environmental-compliance application—align with the results already established for these datasets. We do not reproduce those estimates here. Each application is included to illustrate a different aspect of the model with SEVI errors, and none is intended to be a self-contained replication of the study from which the data are drawn.

7 Conclusions

In a random utility model, individuals choose among alternatives by weighing systematic, observable attributes against a random component known only to the decision-maker. Traditionally, this stochastic element is assumed to follow the right-skewed Largest Extreme Value Type I (LEVI) distribution, leading to the standard multinomial and conditional logit models. This paper challenges that convention by asking a fundamental question: what if the random component instead follows the left-skewed Smallest Extreme Value Type I (SEVI) distribution? *A priori*, there is no theoretical reason to favor LEVI over SEVI beyond a prior belief regarding the nature of unobserved utility shocks.

The SEVI-based choice probabilities possess closed-form expressions that, while more complex than those of the standard logit, remain computationally tractable. With efficient subset enumeration, the SEVI model can be estimated more quickly than the multinomial probit in our timing comparisons, particularly in the moderate range of alternatives common in applied work. Our results show that skewness is a first-order concern for econometricians: in our simulations, incorrectly assuming a LEVI distribution when the true process is SEVI results in substantial misspecification bias in preference parameters and average partial effects, and this bias intensifies as the choice set expands. The two models also assign materially different welfare measures to the same choice environment, as we have shown analytically. Unlike the LEVI-based logit, the SEVI model does not impose IIA. It allows richer substitution patterns without the computational cost of correlated errors.

The SEVI framework provides a foundation for several extensions. Within the utility-maximization framework, the SEVI core can replace the LEVI kernel in mixed logit specifications. In a latent-class context, we may model the populations with heterogeneous skewness types, where one segment of consumers may be characterized by a search for "hidden gems" (LEVI) while another is primarily driven by the avoidance of "hidden flaws" (SEVI). Such a mixed LEVI–SEVI model would treat the direction of unobserved skewness as a latent trait, and it would nest the two specifications compared here as the endpoints of its mixing parameter.

The properties of mixed, nested, and ordered logit models under a SEVI foundation also warrant further investigation. Furthermore, the development of SEVI analogues for rank-ordered data is a natural next step. Section S.F of the supplementary appendix establishes a duality between the LEVI and SEVI choice probabilities and extends it to the Block–Marschak polynomials that govern ranked choice; it implies that the computational burden of estimating models with fully ranked alternatives is essentially the same for the two distributions.

The framework put forth is not specific to economic choice. Where election campaigns are dominated by negative advertising, voters may be more likely to revise their view of a candidate downward than upward, which suggests that a left-skewed random component may be the more natural specification. In biometrics applications, the random component is the latent tolerance distribution whose skewness can reflect unexpected side effects. In multinomial classification, a SEVI analogue of logistic regression, combined with LASSO, is a natural direction for high-dimensional problems.

Finally, well-designed discrete choice experiments can help determine whether skewness is an inherent trait, such as risk aversion, or a response to the choice environment. Studying how skewness varies across demographics, information-gathering styles, and contexts would tell us more about choice behavior. The framework developed here makes the skewness of the random component available as a modeling choice rather than a maintained assumption.

Appendix of Proofs

A.1 Preliminaries

The two lemmas below are used in the proofs of both Theorem 1 and Theorem 2. Throughout, $v_1, \ldots, v_J$ are systematic utilities with scale values $s_k = \exp(v_k)$, the individual subscript i being suppressed. For each $j \in \{1, \ldots, J-1\}$, we split the alternatives into a low and a high group,

$$\mathcal{J}_{\leq j} := \{1, \ldots, j\}, \ \mathcal{J}_{>j} := \{j+1, \ldots, J\}, \ \sigma_{\leq j} := \sum_{k \in \mathcal{J}_{\leq j}} s_k, \ \sigma_{>j} := \sum_{m \in \mathcal{J}_{>j}} s_m,$$

and we write $\{Y \leq j\}$ for the event that the chosen alternative lies in $\mathcal{J}_{\leq j}$. No ordering of $v_1, \ldots, v_J$ is assumed in either lemma.

Lemma A.1 *Let $j \in \{1, \ldots, J-1\}$ and set*

$$F_{\leq j}(t) := \prod_{k \in \mathcal{J}_{\leq j}} \left(1 - e^{-t/s_k}\right), \ t > 0.$$

Then $F_{\leq j}$ is a continuous, strictly increasing bijection from $(0, \infty)$ onto $(0, 1)$. Write $t(\cdot) := F_{\leq j}^{-1}(\cdot)$. Define, for $u \in (0,1)$ and $m \in \mathcal{J}_{>j}$,

$$a_m(u) := 1 - e^{-t(u)/s_m}, \qquad b_m(u) := u^{s_m/\sigma_{\leq j}}. \tag{29}$$

Then:

(i) $P^{\mathrm{SEVI}}(Y \leq j|v) = \int_0^1 \prod_{m \in \mathcal{J}_{>j}} a_m(u)\, du$ and $P^{\mathrm{LEVI}}(Y \leq j|v) = \int_0^1 \prod_{m \in \mathcal{J}_{>j}} b_m(u)\, du$;

(ii) each of a_m and b_m is a continuous, strictly increasing bijection of $(0,1)$ onto itself, with $a_m(0^+) = b_m(0^+) = 0$ and $a_m(1^-) = b_m(1^-) = 1$;

(iii) a_m and b_m are unchanged when every scale value s_r, $r = 1, \ldots, J$, is multiplied by a common positive constant.

Proof of Lemma A.1. Each factor $t \mapsto 1 - e^{-t/s_k}$ is continuous and strictly increasing from 0 to 1, hence so is the finite product $F_{\leq j}$, which establishes the first claim and makes $t(\cdot)$ well defined, continuous and strictly increasing from $(0,1)$ onto $(0,\infty)$.

For Part (i), let $\mathcal{E}_1, \ldots, \mathcal{E}_J$ be iid standard exponential. The SEVI error satisfies $\varepsilon^{\text{SEVI}} \stackrel{d}{=} \ln \mathcal{E}$, so that $e^{U_k} = e^{v_k + \varepsilon_k} = s_k \mathcal{E}_k$. Since the exponential function is strictly increasing, the alternative attaining $\max_k U_k$ is the one attaining $\max_k e^{U_k}$, and therefore

$$\{Y \leq j\} = \left\{ \max_{k \in \mathcal{J}_{\leq j}} s_k \mathcal{E}_k > \max_{m \in \mathcal{J}_{>j}} s_m \mathcal{E}_m \right\}. \tag{30}$$

By independence, $T := \max_{k \in \mathcal{J}_{\leq j}} s_k \mathcal{E}_k$ has distribution function $F_{\leq j}$. Conditioning (30) on $T = t$ and using the independence of $\{\mathcal{E}_m\}_{m \in \mathcal{J}_{>j}}$ from T, we have

$$P^{\text{SEVI}}(Y \leq j | v) = \int_0^\infty \prod_{m \in \mathcal{J}_{>j}} P(s_m \mathcal{E}_m < t)\, dF_{\leq j}(t) = \int_0^\infty \prod_{m \in \mathcal{J}_{>j}} \left(1 - e^{-t/s_m}\right) dF_{\leq j}(t).$$

Note that the integrand is bounded, and $F_{\leq j}$ is absolutely continuous and strictly increasing. So the change of variable $u = F_{\leq j}(t)$ is valid, and this gives

$$P^{\text{SEVI}}(Y \leq j | v) = \int_0^1 \prod_{m \in \mathcal{J}_{>j}} \left(1 - e^{-F_{\leq j}^{-1}(u)/s_m}\right) du$$

which is the first identity of Part (i) with $a_m(u) = 1 - e^{-F_{\leq j}^{-1}(u)/s_m} = 1 - e^{-t/s_m}$ for $t = t(u) = F_{\leq j}^{-1}(u)$. Note that the dependence of $a_m(u)$ and $t(u)$ on j has been suppressed.

For the second result in Part (i), we have $\sum_{m \in \mathcal{J}_{>j}} s_m / \sigma_{\leq j} = \sigma_{>j} / \sigma_{\leq j}$, hence

$$\int_0^1 \prod_{m \in \mathcal{J}_{>j}} u^{s_m/\sigma_{\leq j}} du = \int_0^1 u^{\sigma_{>j}/\sigma_{\leq j}} du = \frac{\sigma_{\leq j}}{\sigma_{\leq j} + \sigma_{>j}} = P^{\text{LEVI}}(Y \leq j | v),$$

where the last equality follows from the logit form summed over $\ell \in \mathcal{J}_{\leq j}$.

For Part (ii), we note that $t(\cdot)$ is a continuous, strictly increasing bijection from $(0,1)$ onto $(0,\infty)$. Since $y \mapsto 1 - e^{-y/s_m}$ is a continuous, strictly increasing bijection from $(0,\infty)$ onto $(0,1)$, the composition $a_m(\cdot)$ is a continuous, strictly increasing bijection of $(0,1)$ onto itself. Its endpoint limits follow from those of $t(\cdot)$. As $u \downarrow 0$, we have $t(u) \downarrow 0$, and hence $a_m(0^+) = 0$, while as $u \uparrow 1$, we have $t(u) \uparrow \infty$ and hence $a_m(1^-) = 1$. The result for $b_m(\cdot)$ is immediate.

For Part (iii), replacing every s_r by λs_r with $\lambda > 0$ replaces $F_{\leq j}(t)$ by $F_{\leq j}(t/\lambda)$ and hence $t(u)$ by $\lambda t(u)$. So the ratio $t(u)/s_m$ is unchanged. In addition, $s_m/\sigma_{\leq j}$ is a ratio of scale values, and it remains the same if every s_r is replaced by λs_r. ■

Lemma A.2 *Let $p \geq 1$, and let each of $a_1(\cdot), b_1(\cdot), \ldots, a_p(\cdot), b_p(\cdot)$ map $(0,1)$ into $[0,1]$ and be nondecreasing. Define*

$$w_\ell(\cdot) := \prod_{r<\ell} a_r(\cdot) \prod_{r>\ell} b_r(\cdot), \quad \ell = 1, \ldots, p,$$

with the convention that an empty product equals 1. Write $a_\ell = a_\ell(u)$, $b_\ell = b_\ell(u)$, and $w_\ell = w_\ell(u)$. Then:

(i) (Telescoping identity) on $(0,1)$*,*

$$\prod_{\ell=1}^{p} a_\ell - \prod_{\ell=1}^{p} b_\ell = \sum_{\ell=1}^{p} w_\ell \left(a_\ell - b_\ell\right). \tag{31}$$

(ii) (Sign inequality) Suppose that for each ℓ *there is* $u_\ell^* \in [0,1]$ *with* $a_\ell - b_\ell \geq 0$ *on* $(0, u_\ell^*)$ *and* $a_\ell - b_\ell \leq 0$ *on* $(u_\ell^*, 1)$*.*[12] *Suppose also that* $\int_0^1 (a_\ell - b_\ell)\, du \leq 0$ *for each* ℓ*. Then* $\int_0^1 w_\ell \left(a_\ell - b_\ell\right) du \leq 0$ *for each* ℓ*, and consequently*

$$\int_0^1 \prod_{\ell=1}^{p} a_\ell du \leq \int_0^1 \prod_{\ell=1}^{p} b_\ell du.$$

(iii) (Strictness) The conclusion of (ii) is strict if, for a single ℓ*, either*

(iii.a) $\int_0^1 (a_\ell - b_\ell)\, du < 0$ *and the left limit of* w_ℓ *at* u_ℓ^* *(the right limit at* 0 *when* $u_\ell^* = 0$*) is strictly positive, or*

(iii.b) $u_\ell^* \in (0,1)$*,* w_ℓ *is strictly increasing on* $(0,1)$*, and* $a_\ell - b_\ell$ *vanishes only at* u_ℓ^**.*

In case (iii.b), the conclusion of (ii) is strict even when $\int_0^1 (a_\ell - b_\ell)\, du = 0$*.*

(iv) (Reversal) Suppose instead that $a_\ell - b_\ell \leq 0$ *on* $(0, u_\ell^*)$ *and* $a_\ell - b_\ell \geq 0$ *on* $(u_\ell^*, 1)$ *for each* ℓ*, and that* $\int_0^1 (a_\ell - b_\ell)\, du \geq 0$ *for each* ℓ*. Then the conclusions of (ii) and (iii) hold with all inequalities reversed, the multiplier* w_ℓ *being replaced throughout by* $\widetilde{w}_\ell := \prod_{r<\ell} b_r \cdot \prod_{r>\ell} a_r$*.*

Proof of Lemma A.2. For Part (i), put $c_\ell := \prod_{r \leq \ell} a_r \cdot \prod_{r>\ell} b_r$ for $\ell = 0, 1, \ldots, p$, so that $c_0 = \prod_r b_r$ and $c_p = \prod_r a_r$. Consecutive terms c_ℓ and $c_{\ell-1}$ differ in a single factor. Indeed, c_ℓ and $c_{\ell-1}$ share $\prod_{r<\ell} a_r \cdot \prod_{r>\ell} b_r = w_\ell$ and carry a_ℓ and b_ℓ respectively in the remaining slot, so that $c_\ell - c_{\ell-1} = w_\ell \left(a_\ell - b_\ell\right)$. Summing over ℓ telescopes to (31).

For Part (ii), fix ℓ and set $\gamma_\ell := \lim_{u \uparrow u_\ell^*} w_\ell(u)$ if $u_\ell^* > 0$ and $\gamma_\ell := w_\ell\left(0^+\right)$ if $u_\ell^* = 0$. The limit exists because w_ℓ is nondecreasing. Also, $0 \leq \gamma_\ell \leq 1$ because w_ℓ is a product of functions with values in $[0,1]$. Monotonicity gives $w_\ell \leq \gamma_\ell$ on $(0, u_\ell^*)$ and $w_\ell \geq \gamma_\ell$ on $(u_\ell^*, 1)$. Hence $w_\ell - \gamma_\ell$ and $a_\ell - b_\ell$ have opposite signs on each of the two intervals, and so $\left(w_\ell - \gamma_\ell\right)\left(a_\ell - b_\ell\right) \leq 0$ on a set of full measure. Therefore

$$\int_0^1 w_\ell \left(a_\ell - b_\ell\right) du = \int_0^1 \left(w_\ell - \gamma_\ell\right)\left(a_\ell - b_\ell\right) du + \gamma_\ell \int_0^1 \left(a_\ell - b_\ell\right) du \leq \gamma_\ell \int_0^1 \left(a_\ell - b_\ell\right) du \leq 0, \tag{32}$$

where the last step holds because $\gamma_\ell \geq 0$. Integrating (31) and summing the p terms leads to the desired result.

For Part (iii), it suffices that one summand of (31) integrate to a strictly negative number, the others being nonpositive by Part (ii). Under (iii.a), $\gamma_\ell > 0$ and the final inequality in (32) is strict. Under (iii.b), w_ℓ is strictly increasing, so $w_\ell - \gamma_\ell < 0$ on $(0, u_\ell^*)$ and $w_\ell - \gamma_\ell > 0$ on $(u_\ell^*, 1)$. Since $a_\ell - b_\ell$ is nonzero off u_ℓ^*, the integrand $\left(w_\ell - \gamma_\ell\right)\left(a_\ell - b_\ell\right)$ is strictly negative except at the single point u_ℓ^*. The first integral in (32) is then strictly negative, and the conclusion follows because $\gamma_\ell \int_0^1 \left(a_\ell - b_\ell\right) du \leq 0$.

Part (iv) follows by applying Parts (i)–(iii) to the interchanged pairs $(\widetilde{a}_\ell, \widetilde{b}_\ell) := (b_\ell, a_\ell)$, for which the crossing property and the sign of the integrals are as required in Part (ii), and whose multiplier is $\widetilde{w}_\ell$. ■

[12] The endpoint values $u_\ell^* \in \{0,1\}$ are allowed, in which case one of the two intervals is empty and the corresponding requirement is vacuous.

A.2 Proof of Theorem 1

Proof of Theorem 1. Set $s_k = \exp(v_k - v_1)$ for $k = 2, \ldots, J$, so that Part (a) is the case $s_k < 1$ for all k, and Part (b) is the case $s_k > 1$ for all k. This normalization is the one of Lemma A.1 with every scale value divided by s_1, which, by Lemma A.1(iii), changes neither a_k nor b_k below.

Apply Lemma A.1 with the splitting index $j = 1$, so that $\mathcal{J}_{\leq 1} = \{1\}$, $\sigma_{\leq 1} = s_1 = 1$ and $\{Y \leq 1\} = \{Y = 1\}$. Here the low group is the single alternative 1, so $F_{\leq 1}(t) = 1 - e^{-t}$, which inverts explicitly to $t(u) = -\ln(1-u)$. Substituting this into (29) gives the closed forms

$$a_k(u) = 1 - (1-u)^{1/s_k}, \qquad b_k(u) = u^{s_k}, \qquad k = 2, \ldots, J, \tag{33}$$

and Lemma A.1(i) gives

$$P^{\text{SEVI}}(Y = 1|v) = \int_0^1 \prod_{k=2}^{J} a_k(u)du, \qquad P^{\text{LEVI}}(Y = 1|v) = \frac{1}{1 + \sum_{k=2}^{J} s_k} = \int_0^1 \prod_{k=2}^{J} b_k(u)du.$$

By Lemma A.1(ii), each a_k and each b_k is continuous and strictly increasing on $(0,1)$, with limits 0 at $u = 0$ and 1 at $u = 1$. Noting that $\int_0^1 (1-u)^{1/s_k} du = s_k/(1+s_k)$ and $\int_0^1 u^{s_k} du = 1/(1+s_k)$, we have

$$\int_0^1 a_k(u)du = \int_0^1 b_k(u)du = \frac{1}{1+s_k}, \qquad \text{so} \qquad \int_0^1 [a_k(u) - b_k(u)]\, du = 0. \tag{34}$$

We next show that, for each k, the function $a_k(u) - b_k(u)$ changes sign exactly once over $u \in (0,1)$. We consider a fixed k and suppress the dependence on k for notational economy. Under either Part (a) or Part (b), $s_k \neq 1$, so we write $s = s_k$ and let

$$\Delta(u) = a_k(u) - b_k(u) = 1 - (1-u)^{1/s} - u^s.$$

Then, $\Delta(0^+) = \Delta(1^-) = 0$, and the derivative of $\Delta(u)$ is

$$\Delta'(u) = \frac{1}{s}(1-u)^{1/s-1} - su^{s-1}.$$

Now, $\Delta'(u) = 0$ if and only if $(1-u)^{1/s-1} = s^2 u^{s-1}$. Since both sides are positive, taking logarithms shows that $\Delta'(u) = 0$ if and only if $\kappa(u) = 0$, where

$$\kappa(u) = \frac{1}{s}\ln(1-u) + \ln u - \frac{2\ln s}{1-s}.$$

Note that

$$\kappa'(u) = \frac{1}{u} - \frac{1}{s(1-u)} = \frac{(1+s)}{su(1-u)}\left(\frac{s}{1+s} - u\right),$$

which is positive on $(0, s/(1+s))$ and negative on $(s/(1+s), 1)$. So, $\kappa(u)$ is strictly increasing on $(0, s/(1+s))$ and strictly decreasing on $(s/(1+s), 1)$, with $\kappa(0^+) = \kappa(1^-) = -\infty$. Hence, $\kappa(u)$ has at most two zeros, which implies that $\Delta'(\cdot)$ also has at most two zeros on $(0,1)$.

Consider the case that $s \in (0,1)$ so that $1/s - 1 > 0$ and $s - 1 < 0$. For illustration, Figure 8 plots the functions $\Delta'(u)$ and $\Delta(u)$ when $s = 1/2$. For any $s \in (0,1)$, we have $\Delta'(0^+) = -\infty$ and $\Delta'(1^-) = -s < 0$, so $\Delta'(\cdot)$ is negative near both endpoints. Recall that $\Delta'(\cdot)$ has at most two zeros on $(0,1)$. We next show that $\Delta'(\cdot)$ has exactly two zeros on $(0,1)$. We use the

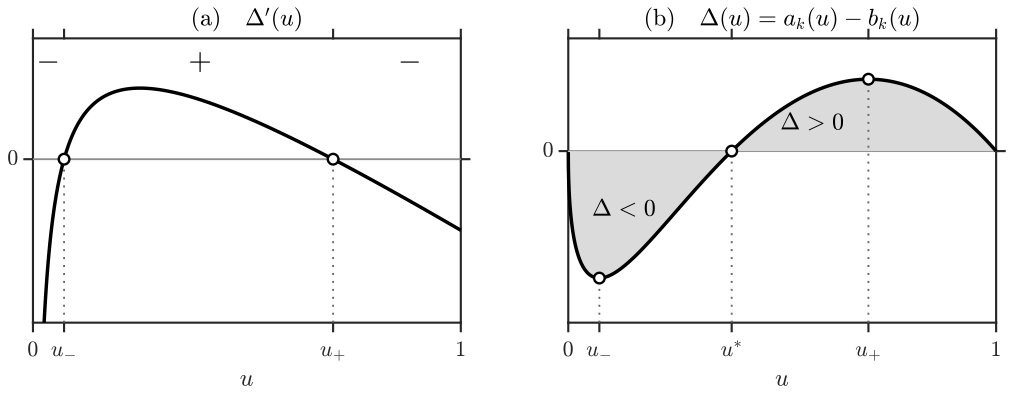


Figure 8: The two steps of the single-crossing argument, shown for $s = 1/2$. Panel (a): Δ' has exactly two zeros $u_- < u_+$ on $(0,1)$, with sign pattern $-$, $+$, $-$. Panel (b): $\Delta = a_k - b_k$ therefore crosses zero exactly once, at u^*, and attains its minimum at u_- and its maximum at u_+. For $s > 1$, the signs in both panels are reversed.

following observation repeatedly: for a differentiable function $f(\cdot)$ on $(0,1)$ with derivative $f'(\cdot)$, if $f'(u) \leq 0$ on $(0,1)$ with equality at finitely many points (possibly none), then $f(\cdot)$ is strictly decreasing on $(0,1)$. The observation can be easily proved using the fundamental theorem of calculus, and it helps us rule out the first two cases below.

Case 1: $\Delta'(\cdot)$ has no zero. Being continuous and negative near the endpoints, $\Delta'(\cdot)$ is then negative on all of $(0,1)$. As a result, $\Delta(\cdot)$ is strictly decreasing on $(0,1)$, which contradicts $\Delta(0^+) = \Delta(1^-) = 0$.

Case 2: $\Delta'(\cdot)$ has a single zero u_o at which it does not change sign. Then $\Delta'(u) \leq 0$ on $(0,1)$, with equality only at u_o. So $\Delta(\cdot)$ is strictly decreasing on $(0,1)$, contradicting $\Delta(0^+) = \Delta(1^-) = 0$.

Case 3: $\Delta'(\cdot)$ has a single zero u_o at which it changes sign. Then $\Delta'(\cdot)$ takes opposite signs on $(0, u_o)$ and $(u_o, 1)$, contradicting that $\Delta'(\cdot)$ is negative near both endpoints.

Therefore, none of the above three cases is possible, and $\Delta'(\cdot)$ has exactly two zeros $u_- < u_+$ on $(0,1)$, and it has a constant sign on each of the three intervals $(0, u_-)$, (u_-, u_+), and $(u_+, 1)$. That sign is negative on the first and third intervals. It is positive on the second, since otherwise $\Delta'(u) \leq 0$ on $(0,1)$ with equality only at u_- and u_+, and then Δ would be strictly decreasing on $(0,1)$, contradicting $\Delta(0^+) = \Delta(1^-) = 0$. The sign pattern is therefore $-$, $+$, $-$. Consequently, $\Delta(\cdot) < 0$ on $(0, u_-]$, $\Delta(\cdot) > 0$ on $[u_+, 1)$, and $\Delta(\cdot)$ increases strictly on (u_-, u_+), so it vanishes at exactly one interior point u^* and

$$\Delta < 0 \text{ on } (0, u^*), \qquad \Delta > 0 \text{ on } (u^*, 1).$$

The case $s > 1$ is symmetric and is treated in Part (b) below. In that case, $\Delta(\cdot)$ has a unique interior zero, which we again denote by u^*.

To combine these facts, we apply Lemma A.2 to the $J-1$ pairs $(a_k, b_k)_{k=2}^J$, which map $(0,1)$ into $[0,1]$ and are nondecreasing by Lemma A.1(ii). The multiplier appearing in that lemma is

$$w_k := \prod_{m=2}^{k-1} a_m \cdot \prod_{m=k+1}^{J} b_m \text{ or } \widetilde{w}_k := \prod_{m=2}^{k-1} b_m \cdot \prod_{m=k+1}^{J} a_m.$$

Since $J > 2$, each of w_k and $\widetilde{w}_k$ is a product of $J - 2 \geq 1$ strictly increasing functions with values in $(0,1)$, so both are strictly positive and strictly increasing on $(0,1)$. For each k, let u_k^* denote

the crossing point u^* constructed above with $s = s_k$. It lies in $(0,1)$, and $a_k - b_k$ vanishes nowhere else on that interval.

In Part (a), we have $s_k < 1$ for every k. We have shown that $a_k - b_k < 0$ on $(0, u_k^*)$ and $a_k - b_k > 0$ on $(u_k^*, 1)$, while (34) gives $\int_0^1 (a_k - b_k)\, du = 0$. This is the configuration of Lemma A.2(iv), whose condition (iii.b) is met because $\widetilde{w}_k$ is strictly increasing and $a_k - b_k$ vanishes only at the interior point u_k^*. Hence

$$P^{\text{SEVI}}(Y = 1|v) = \int_0^1 \prod_{k=2}^{J} a_k du > \int_0^1 \prod_{k=2}^{J} b_k du = P^{\text{LEVI}}(Y = 1|v),$$

which proves Part (a). The inequality is strict even though $a_k - b_k$ integrates to zero. Strictness comes from the strict monotonicity of the multiplier, which is condition (iii.b) in Lemma A.2.

For Part (b) with $s > 1$, we have $\Delta'(0^+) = 1/s > 0$ and $\Delta'(1^-) = +\infty$. The same argument, with the inequalities reversed, shows that the sign pattern of $\Delta'(\cdot)$ is $+, -, +$ and gives $\Delta(u) > 0$ on $(0, u^*)$ and $\Delta(u) < 0$ on $(u^*, 1)$. Since $u_k^* = u^*$ for $s = s_k$, the crossing of $a_k - b_k$ is now in the orientation assumed in Lemma A.2(ii) and (iii). Again $\int_0^1 (a_k - b_k)\, du = 0$ and w_k is strictly increasing, so Lemma A.2(iii.b) applies and yields the reverse strict inequality. Part (b) follows.

For Part (c) we return to the original scale values $s_k = \exp(v_k)$ and write $r_k := s_{1|k} = s_1/s_k$ for $k = 2, \ldots, J$, so that every $r_k \to 0$ as $s_1 \to 0$ with $v_2, \ldots, v_J$ held fixed. Substituting $u = e^{-t}$ in (3) gives

$$P^{\text{SEVI}}(Y = 1|v) = \int_0^\infty e^{-t} \prod_{k=2}^{J} \left(1 - e^{-r_k t}\right) dt,$$

and hence

$$\frac{P^{\text{SEVI}}(Y = 1|v)}{\prod_{k=2}^{J} r_k} = \int_0^\infty e^{-t} \prod_{k=2}^{J} \frac{1 - e^{-r_k t}}{r_k} dt.$$

For each fixed $t > 0$, every factor of the product converges to t as $r_k \downarrow 0$, and $1 - e^{-x} \leq x$ bounds each factor by t, so the integrand is dominated by $t^{J-1} e^{-t}$, which is integrable on $(0, \infty)$. Dominated convergence therefore gives

$$\frac{P^{\text{SEVI}}(Y = 1|v)}{\prod_{k=2}^{J} r_k} \longrightarrow \int_0^\infty t^{J-1} e^{-t} dt = (J-1)!.$$

Since $\prod_{k=2}^{J} r_k = s_1^{J-1} \prod_{k=2}^{J} s_k^{-1}$, this is the first line of (7). For the second, (5) gives $P^{\text{LEVI}}(Y = 1|v) = s_1/(s_1 + \sum_{k=2}^{J} s_k)$, and the denominator tends to $\sum_{k=2}^{J} s_k$ as $s_1 \to 0$. Part (c) follows. ■

A.3 Proofs of Theorem 2 and Corollary 1

The proof of Theorem 2 uses Lemmas A.1 and A.2 together with three further results. Throughout, $v_1 \leq \cdots \leq v_J$, the index $j \in \{1, \ldots, J-1\}$ is fixed, and the functions a_m and b_m are as in (29), and their dependence on j has been suppressed. For $x > 0$, we write $\mathcal{Y}(x) := -\ln(1 - e^{-x})$. This function is positive, strictly decreasing and real-analytic on $(0, \infty)$, with $\mathcal{Y}(0^+) = \infty$, $\mathcal{Y}(\infty) = 0$ and $\mathcal{Y}'(x) = -1/(e^x - 1)$.

Lemma A.3 *The function $x \mapsto \mathcal{Y}(\zeta x)/\mathcal{Y}(x)$ is nonincreasing on $(0, \infty)$ for every $\zeta \geq 1$, and strictly decreasing for every $\zeta > 1$.*

Proof of Lemma A.3. Write $\rho(x) := x\mathcal{Y}'(x)/\mathcal{Y}(x)$ for the elasticity of $\mathcal{Y}$. Since $\mathcal{Y} > 0$, we may differentiate the logarithm of the ratio:

$$\frac{d}{dx}\ln\frac{\mathcal{Y}(\zeta x)}{\mathcal{Y}(x)} = \zeta\frac{\mathcal{Y}'(\zeta x)}{\mathcal{Y}(\zeta x)} - \frac{\mathcal{Y}'(x)}{\mathcal{Y}(x)} = \frac{1}{x}\left[\rho(\zeta x) - \rho(x)\right]. \tag{35}$$

The ratio $\mathcal{Y}(\zeta x)/\mathcal{Y}(x)$ is positive, so its derivative has the same sign as the right-hand side of (35). Since $\zeta x \geq x$, with strict inequality when $\zeta > 1$, it therefore suffices to show that ρ is strictly decreasing on $(0,\infty)$.

We next prove that

$$\mathcal{Y}(x)\frac{xe^x - e^x + 1}{x} < 1 \text{ for every } x > 0 \tag{36}$$

using the following argument. Since $xe^x - e^x + 1 = 0$ when $x = 0$, and

$$\frac{d}{dx}\left(xe^x - e^x + 1\right) = xe^x > 0,$$

we have $xe^x - e^x + 1 > 0$ for $x > 0$. Next, noting that

$$\mathcal{Y}(x) = \ln\left(1 + \frac{1}{(e^x - 1)}\right) < \frac{1}{e^x - 1},$$

and multiplying by the positive quantity $(xe^x - e^x + 1)/x$, we obtain

$$\mathcal{Y}(x)\frac{xe^x - e^x + 1}{x} < \frac{xe^x - e^x + 1}{x(e^x - 1)}.$$

The above upper bound is at most 1, because

$$\frac{xe^x - e^x + 1}{x(e^x - 1)} \leq 1 \Longleftrightarrow xe^x - e^x + 1 \leq xe^x - x \Longleftrightarrow x \leq e^x - 1,$$

the last inequality holding for every $x \geq 0$. This proves (36).

It remains to deduce that ρ is strictly decreasing. The function $\mathcal{Y}$ is smooth and positive, so ρ is differentiable with

$$\rho'(x) = \frac{(\mathcal{Y}' + x\mathcal{Y}'')\mathcal{Y} - x(\mathcal{Y}')^2}{\mathcal{Y}^2}.$$

Given that $\mathcal{Y}' = -1/(e^x - 1)$ and $\mathcal{Y}'' = e^x/(e^x - 1)^2$, we have

$$\mathcal{Y}'(x) + x\mathcal{Y}''(x) = \frac{xe^x - (e^x - 1)}{(e^x - 1)^2}, \qquad x\left[\mathcal{Y}'(x)\right]^2 = \frac{x}{(e^x - 1)^2}.$$

Hence,

$$\rho'(x) = \frac{\left[xe^x - e^x + 1\right]\mathcal{Y}(x) - x}{(e^x - 1)^2\mathcal{Y}(x)^2}.$$

By (36), the numerator is strictly negative, so $\rho'(x) < 0$ on $(0,\infty)$ and ρ is strictly decreasing on $(0,\infty)$. ■

Lemma A.4 *Suppose $v_j < v_{j+1}$. Then for every $m \in \mathcal{J}_{>j}$ there is $u_m^* \in (0,1)$ such that $a_m - b_m > 0$ on $(0, u_m^*)$ and $a_m - b_m < 0$ on $(u_m^*, 1)$.*

Proof of Lemma A.4. Fix $m \in \mathcal{J}_{>j}$. Since v is nondecreasing and $v_j < v_{j+1}$, we have $v_k \le v_j < v_{j+1} \le v_m$ for every $k \in \mathcal{J}_{\le j}$, so that

$$s_k < s_m \text{ for every } k \in \mathcal{J}_{\le j}. \tag{37}$$

Write $\theta := s_m/\sigma_{\le j}$, so that $b_m = u^\theta$ by (29). Throughout the proof we write $t = t(u) = F_{\le j}^{-1}(u)$, which is the change of variable used in Lemma A.1, so that $u = F_{\le j}(t)$, the definition of a_m in (29) reads $a_m(u) = 1 - e^{-t/s_m}$, and t ranges over $(0,\infty)$ as u ranges over $(0,1)$. Because j is now arbitrary, $t(\cdot)$ is not available in closed form, and a_m cannot be reduced to an explicit function of u as it was in (33). For $u \in (0,1)$, let

$$\mathcal{R}(u) := \frac{\ln a_m(u)}{\ln u} = \frac{\ln\left(1 - e^{-t(u)/s_m}\right)}{\ln u}.$$

Both $a_m(u)$ and u lie in $(0,1)$, so $\ln a_m(u) < 0$ and $\ln u < 0$. The logarithm is strictly increasing, so each of the relations between a_m and b_m is equivalent to the same relation between $\ln a_m$ and $\ln b_m = \theta \ln u$. Dividing by $\ln u < 0$ reverses the two strict inequalities and preserves the equality, so that

$$a_m > b_m \Longleftrightarrow \mathcal{R} < \theta,\ a_m < b_m \Longleftrightarrow \mathcal{R} > \theta,\ a_m = b_m \Longleftrightarrow \mathcal{R} = \theta. \tag{38}$$

In terms of the variable t, we have $\ln a_m = \ln\left(1 - e^{-t/s_m}\right) = -\mathcal{Y}(t/s_m)$, while $u = F_{\le j}(t)$ gives $\ln u = \sum_{k\in\mathcal{J}_{\le j}} \ln\left(1 - e^{-t/s_k}\right) = -\sum_{k\in\mathcal{J}_{\le j}} \mathcal{Y}(t/s_k)$. Hence

$$\mathcal{R} = \frac{\mathcal{Y}(t/s_m)}{\sum_{k\in\mathcal{J}_{\le j}} \mathcal{Y}(t/s_k)}. \tag{39}$$

By (37), each $k \in \mathcal{J}_{\le j}$ has $\zeta_k := s_m/s_k > 1$, and $\mathcal{Y}(t/s_k) = \mathcal{Y}(\zeta_k \cdot t/s_m)$. Applying Lemma A.3 with $\zeta = \zeta_k$ and $x = t/s_m$, each function $t \mapsto \mathcal{Y}(t/s_k)/\mathcal{Y}(t/s_m)$ is strictly decreasing. Dividing the numerator and the denominator of (39) by $\mathcal{Y}(t/s_m)$ gives

$$\mathcal{R} = \left[\sum_{k\in\mathcal{J}_{\le j}} \frac{\mathcal{Y}(t/s_k)}{\mathcal{Y}(t/s_m)}\right]^{-1}.$$

The sum in the bracket is a finite sum of positive, strictly decreasing functions of t, hence is itself positive and strictly decreasing. Its reciprocal is therefore strictly increasing, so $\mathcal{R}$ is a strictly increasing function of t, and hence of u.

We compute its limits. As $y \downarrow 0$, we have $(1 - e^{-y}) = y(1 + o(1))$, so $\mathcal{Y}(y) = (-\ln y)(1 + o(1))$. It then follows that

$$\lim_{u\to 0+} \mathcal{R}(u) = \lim_{t\to 0+} \frac{-\ln t + \ln s_m}{-j \ln t + \sum_{k\in\mathcal{J}_{\le j}} \ln s_k} = \frac{1}{j}.$$

As $y \to \infty$, $\mathcal{Y}(y) = e^{-y}(1 + o(1))$. By (37), $e^{-t/s_k} = o\left(e^{-t/s_m}\right)$ as $t \to \infty$ for every $k \in \mathcal{J}_{\le j}$. Summing over the j members of $\mathcal{J}_{\le j}$, the denominator of (39) is $o\left(e^{-t/s_m}\right)$ while its numerator is $e^{-t/s_m}(1 + o(1))$, so $\mathcal{R}(u) \to \infty$ as $u \to 1-$. Finally, (37) gives $\sigma_{\le j} = \sum_{k\in\mathcal{J}_{\le j}} s_k < j s_m$, so $\theta > 1/j$. Since $\mathcal{R}$ is continuous and strictly increasing from $1/j$ to ∞, there is a unique $u_m^* \in (0,1)$ with $\mathcal{R}(u_m^*) = \theta$, and $\mathcal{R} < \theta$ on $(0, u_m^*)$, $\mathcal{R} > \theta$ on $(u_m^*, 1)$. By (38), this is the assertion. ■

Lemma A.5 *Suppose $v_j < v_{j+1}$. Then for every $m \in \mathcal{J}_{>j}$,*

$$\int_0^1 [a_m(u) - b_m(u)]\, du \leq 0,$$

with strict inequality when $j \geq 2$.

Proof of Lemma A.5. Apply Lemma A.1 to the auxiliary model with $j+1$ alternatives whose systematic utilities are $(v_k)_{k\in\mathcal{J}_{\leq j}}$ together with v_m, split into the low group $\mathcal{J}_{\leq j}$ and the high group $\{m\}$. By (29), the functions a_m and b_m depend only on $\mathcal{J}_{\leq j}$ and on m, so they are unchanged by passing to this smaller model, and Lemma A.1(i) gives

$$\int_0^1 a_m(u)\, du = P^{\mathrm{SEVI}}(Y \in \mathcal{J}_{\leq j}), \qquad \int_0^1 b_m(u)\, du = \frac{\sigma_{\leq j}}{\sigma_{\leq j} + s_m} = P^{\mathrm{LEVI}}(Y \in \mathcal{J}_{\leq j}),$$

both probabilities referring to the auxiliary model with $j+1$ alternatives.

Since $m \in \mathcal{J}_{>j}$ and $v_j < v_{j+1}$, we have $v_m \geq v_{j+1} > v_j \geq v_k$ for every $k \in \mathcal{J}_{\leq j}$, so alternative m has the strictly largest scale among the $(j+1)$ alternatives in the auxiliary model. If $j \geq 2$, the auxiliary model has $j+1 \geq 3$ alternatives, so Theorem 1(a) applies and gives $P^{\mathrm{SEVI}}(Y = m) > P^{\mathrm{LEVI}}(Y = m)$. If $j = 1$, the auxiliary model has two alternatives, for which the SEVI and LEVI specifications coincide and the two probabilities are equal. Taking complements within the auxiliary model yields the assertion in both cases. ■

Proof of Theorem 2. We prove Part (a) first, then Part (b).

Part (a). Fix $j \in \{1, \ldots, J-1\}$ and suppose first that $v_j < v_{j+1}$. By Lemma A.1, the two probabilities are the integrals over $(0,1)$ of the products of a_m and of b_m over $m \in \mathcal{J}_{>j}$. By Lemma A.1(ii), these factors map $(0,1)$ into $[0,1]$ and are nondecreasing, so Lemma A.2 applies to them. Its crossing condition is supplied by Lemma A.4 and its integral condition by Lemma A.5, both of which apply because $v_j < v_{j+1}$. Lemma A.2(ii) therefore gives

$$P^{\mathrm{SEVI}}(Y \leq j|v) = \int_0^1 \prod_{m\in\mathcal{J}_{>j}} a_m du \leq \int_0^1 \prod_{m\in\mathcal{J}_{>j}} b_m du = P^{\mathrm{LEVI}}(Y \leq j|v). \tag{40}$$

Suppose next that $v_j = v_{j+1}$. For $\varepsilon > 0$, define $v^{(\varepsilon)}$ by $v_r^{(\varepsilon)} := v_r$ for $r \leq j$ and $v_r^{(\varepsilon)} := v_r + \varepsilon$ for $r > j$. This vector is again nondecreasing and satisfies $v_j^{(\varepsilon)} < v_{j+1}^{(\varepsilon)}$, so (40) holds for $v^{(\varepsilon)}$ at the index j. Both sides of (40) are continuous in v: for LEVI this is immediate from $P^{\mathrm{LEVI}}(Y \leq j|v) = \sigma_{\leq j}/(\sigma_{\leq j} + \sigma_{>j})$, while for SEVI, (30) expresses the probability as $P(\max_{k\in\mathcal{J}_{\leq j}} s_k\mathcal{E}_k > \max_{m\in\mathcal{J}_{>j}} s_m\mathcal{E}_m)$, whose indicator is, for almost every realization of $\mathcal{E}$, continuous at the point s: ties between the two maxima occur with probability zero, and in their absence the indicator is locally constant in s. Dominated convergence applies, and letting $\varepsilon \downarrow 0$ gives (40) at v.

Part (b). Assume $v_1 < v_J$ and write

$$\Xi(\ell) := P^{\mathrm{LEVI}}(Y \leq \ell|v) - P^{\mathrm{SEVI}}(Y \leq \ell|v), \ \ell = 0, 1, \ldots, J,$$

so that $\Xi(0) = \Xi(J) = 0$ and $\Xi \geq 0$ on $\{1, \ldots, J-1\}$ by Part (a). Its increments are

$$\Xi(\ell) - \Xi(\ell-1) = P^{\mathrm{LEVI}}(Y = \ell|v) - P^{\mathrm{SEVI}}(Y = \ell|v). \tag{41}$$

Fix again an index $j \in \{1, \ldots, J-1\}$. We show that $\Xi(j) > 0$.

Case 1: $v_j < v_{j+1}$. Suppose first $j \geq 2$. Then Lemma A.5 gives $\int_0^1 (a_m - b_m)\, du < 0$ for every $m \in \mathcal{J}_{>j}$, and Lemma A.4 places each crossing at an interior point $u_m^* \in (0,1)$. The multiplier w_m of Lemma A.2 is a product of functions that are strictly positive on $(0,1)$, and $w_m \equiv 1$ when $\mathcal{J}_{>j} = \{m\}$. In either case $w_m > 0$ on $(0,1)$, and w_m is nondecreasing, so its left limit at the interior point u_m^* is at least $w_m(u) > 0$ for any $u < u_m^*$, hence strictly positive. Condition (iii.a) in Lemma A.2 therefore applies, and the inequality (40) is strict. That is, $\Xi(j) > 0$. Suppose instead $j = 1$. Then $v_1 < v_r$ for every $r > 1$, and $\Xi(1) > 0$ is exactly Theorem 1(b).

Case 2: $v_j = v_{j+1}$. Let $B = \{p, \ldots, q\}$ be the maximal block of indices on which v is constant and which contains both j and $j+1$, so that $p \leq j < q$. The alternatives in B share a systematic utility, hence are exchangeable under either specification and carry equal choice probabilities. By (41), the increment $\Xi(\ell) - \Xi(\ell - 1)$ is then the same for every $\ell \in B$, so Ξ is affine on $\{p-1, p, \ldots, q\}$. Consequently

$$\Xi(j) = \frac{q-j}{q-p+1}\Xi(p-1) + \frac{j-p+1}{q-p+1}\Xi(q), \tag{42}$$

a convex combination whose two weights are strictly positive because $p \leq j \leq q-1$. Both $\Xi(p-1)$ and $\Xi(q)$ are nonnegative: by Part (a) when the index lies in $\{1, \ldots, J-1\}$, and because $\Xi(0) = \Xi(J) = 0$ otherwise. If $p > 1$, then $v_{p-1} < v_p$, so $p-1$ is an index of Case 1 and $\Xi(p-1) > 0$. If $q < J$, then $v_q < v_{q+1}$, so q is an index of Case 1 and $\Xi(q) > 0$. Since $v_1 < v_J$, the block B is a proper subset of $\{1, \ldots, J\}$, so $p > 1$ or $q < J$, and (42) gives $\Xi(j) > 0$. ■

Proof of Corollary 1. Part (a): writing $F_j^{\mathcal{M}} := P^{\mathcal{M}}(Y \leq j|v)$, Abel summation gives

$$\left\langle P^{\text{SEVI}}, w \right\rangle - \left\langle P^{\text{LEVI}}, w \right\rangle = \sum_{j=1}^{J-1} \left(F_j^{\text{LEVI}} - F_j^{\text{SEVI}}\right)(w_{j+1} - w_j).$$

Every bracket is nonnegative by Theorem 2(a) and every increment is nonnegative because w is nondecreasing, so the sum is nonnegative; this is the standard characterization of first-order stochastic dominance on an ordered finite set. If $v_1 < v_J$, Theorem 2(b) makes every bracket strictly positive, so the sum is strictly positive as soon as one increment is. Part (b) is Part (a) at $w = v$, which is nondecreasing by assumption, and $v_1 < v_J$ forces $v_{j+1} > v_j$ for at least one j.

Part (c) needs one further fact, that $P^{\mathcal{M}}(v)$ is itself nondecreasing. By (3), $P^{\text{SEVI}}(Y = j|v) = \int_0^1 \prod_{m \neq j} \left[1 - u^{s_j/s_m}\right] du$ is strictly increasing in s_j and strictly decreasing in each s_m, and the same is immediate for (5). Suppose $v_j < v_k$ and let v' exchange those two coordinates. The random components are iid, so $P^{\mathcal{M}}(Y = j|v') = P^{\mathcal{M}}(Y = k|v)$, while passing from v to v' raises the j-th coordinate and lowers the k-th and therefore raises $P^{\mathcal{M}}(Y = j|\cdot)$; hence $P^{\mathcal{M}}(Y = k|v) > P^{\mathcal{M}}(Y = j|v)$. Applying Part (a) at $w = P^{\text{SEVI}}(v)$ and at $w = P^{\text{LEVI}}(v)$ and chaining the two,

$$\left\langle P^{\text{SEVI}}, P^{\text{SEVI}} \right\rangle \geq \left\langle P^{\text{LEVI}}, P^{\text{SEVI}} \right\rangle = \left\langle P^{\text{SEVI}}, P^{\text{LEVI}} \right\rangle \geq \left\langle P^{\text{LEVI}}, P^{\text{LEVI}} \right\rangle.$$

If $v_1 < v_J$, the monotonicity just established gives $P^{\text{SEVI}}(Y = J|v) > P^{\text{SEVI}}(Y = 1|v)$, so $w = P^{\text{SEVI}}(v)$ has at least one strict increment and the first inequality in the chain is strict by Part (a). ■

Supplementary Appendix[13]

This appendix collects material that supports the main text. Section S.A proves Lemmas 1 and 2; Sections S.B–S.D prove Theorem 3 and Propositions 1, 2, and 3; and Section S.E proves Theorem 4 on the consistency of the MLE and Theorem 5 on the estimation of parameter ratios. Section S.F presents the duality between the SEVI and LEVI choice probabilities and uses the duality to characterize the Block–Marschak polynomials. Section S.G sharpens the "Leader-Takes-More" effect of Theorem 1: Section S.G.1 characterizes the sign of the probability difference when $J = 3$, where the alternative under consideration may not have the highest or lowest systematic utility, and Section S.G.2 traces the effect to the largest advantage that any rival draws over a fixed alternative. Section S.H reports the designs and the numerical results behind the choice probability, compensating variation, and computation time comparisons summarized in Sections 2, 3.3, and 5.3. Section S.I collects four further figures from the empirical applications.

S.A Proofs of Lemmas 1 and 2

The derivations below put Lemmas 1 and 2 in the notation used in the main text and, for Lemma 1, in the integral form on which the proofs in Appendix A operate.

Proof of Lemma 1. Let $G(\cdot)$ and $g(\cdot)$ denote the cumulative distribution function (CDF) and probability density function (pdf) of the standard Gumbel (LEVI) distribution:

$$G(z) = \exp(-e^{-z}), \quad g(z) = e^{-z}\exp(-e^{-z}). \tag{S.1}$$

Let $F(\cdot)$ and $f(\cdot)$ be the CDF and pdf of the SEVI distribution. Since the SEVI distribution is the reflection of the Gumbel distribution, we have $F(\varepsilon) = 1 - G(-\varepsilon)$ and $f(\varepsilon) = g(-\varepsilon)$.

Since ε_i is independent of V_i and ε_{ij} is iid over j, we first condition on $\varepsilon_{ij} = z$ and then integrate z out using the pdf of ε_{ij}:

$$P^{\text{SEVI}}(Y_i = j|v_i) = \int_{-\infty}^{\infty} \left(\prod_{k \in \mathcal{J}_{-j}} \Pr(\varepsilon_{ik} \le v_{ij} - v_{ik} + z) \right) f(z)dz.$$

Substituting the reflection properties of the SEVI distribution yields:

$$\begin{aligned} P^{\text{SEVI}}(Y_i = j|v_i) &= \int_{-\infty}^{\infty} \left(\prod_{k \in \mathcal{J}_{-j}} [1 - G(v_{ik} - v_{ij} - z)] \right) g(-z)dz \\ &= \int_{-\infty}^{\infty} \left(\prod_{k \in \mathcal{J}_{-j}} [1 - G(v_{ik} - v_{ij} + w)] \right) g(w)dw. \end{aligned}$$

Note that $v_{ik} - v_{ij} = \ln s^i_{k|j}$. Setting $u = G(w) \in (0, 1)$ so that $du = g(w)dw$, we have

$$\begin{aligned} G(v_{ik} - v_{ij} + w) &= G(\ln s^i_{k|j} + w) = \exp(-e^{-(\ln s^i_{k|j} + w)}) = \exp(-e^{-w}/s^i_{k|j}) \\ &= \exp((\ln u)/s^i_{k|j}) = \exp(\ln u^{1/s^i_{k|j}}) = u^{1/s^i_{k|j}}. \end{aligned}$$

[13] Due to space constraints, this supplementary appendix is not intended for publication; it will be made available online.

Substituting these into the expression for $P^{\text{SEVI}}(Y_i = j|v_i)$ yields:

$$P^{\text{SEVI}}(Y_i = j|v_i) = \int_0^1 \prod_{k \in \mathcal{J}_{-j}} \left[1 - u^{1/s^i_{k|j}}\right] du. \tag{S.2}$$

Since $1/s^i_{k|j} = s^i_{j|k}$, this is (3). Expanding the product by inclusion-exclusion and integrating term by term gives (4):

$$\begin{aligned} P^{\text{SEVI}}(Y_i = j|v_i) &= 1 + \sum_{\ell=1}^{J-1} (-1)^\ell \sum_{|\mathcal{S}|=\ell, \mathcal{S} \subseteq \mathcal{J}_{-j}} \int_0^1 u^{\sum_{k \in \mathcal{S}} s^i_{j|k}} du \\ &= 1 + \sum_{\ell=1}^{J-1} (-1)^\ell \sum_{|\mathcal{S}|=\ell, \mathcal{S} \subseteq \mathcal{J}_{-j}} \frac{1}{1 + \sum_{k \in \mathcal{S}} s^i_{j|k}}. \end{aligned}$$

■

Proof of Lemma 2. Recall that $G(\cdot)$ and $g(\cdot)$ are the CDF and the pdf of the LEVI distribution. For $\varepsilon_j \sim iid$ SEVI and $\tilde{\varepsilon}_j \sim iid$ LEVI, we have

$$\begin{aligned} & E\left[\max_{j=1,\ldots,J} \{v_j + \varepsilon_j\}\right] \\ &= \sum_{j=1}^J E\left[v_j + \varepsilon_j | v_j + \varepsilon_j \geq v_k + \varepsilon_k,\ k \in \mathcal{J}_{-j}\right] P^{\text{SEVI}}(Y = j|v) \\ &= \sum_{j=1}^J E\left[v_j - \tilde{\varepsilon}_j | v_j - \tilde{\varepsilon}_j \geq v_k - \tilde{\varepsilon}_k,\ k \in \mathcal{J}_{-j}\right] P^{\text{SEVI}}(Y = j|v) \\ &= \sum_{j=1}^J E\left[v_j - \tilde{\varepsilon}_j | - v_j + \tilde{\varepsilon}_j \leq -v_k + \tilde{\varepsilon}_k,\ k \in \mathcal{J}_{-j}\right] P^{\text{SEVI}}(Y = j|v) \\ &= \sum_{j=1}^J \left[v_j - E\left(\tilde{\varepsilon}_j | v_k - v_j + \tilde{\varepsilon}_j \leq \tilde{\varepsilon}_k,\ k \in \mathcal{J}_{-j}\right)\right] P^{\text{SEVI}}(Y = j|v) \\ &= \sum_{j=1}^J v_j P^{\text{SEVI}}(Y = j|v) \\ &- \sum_{j=1}^J E\left(\tilde{\varepsilon}_j | v_k - v_j + \tilde{\varepsilon}_j \leq \tilde{\varepsilon}_k,\ k \in \mathcal{J}_{-j}\right) P^{\text{SEVI}}(Y = j|v). \end{aligned}$$

Next,

$$
\begin{aligned}
&E\left(\tilde{\varepsilon}_j | v_k - v_j + \tilde{\varepsilon}_j \leq \tilde{\varepsilon}_k,\ k \in \mathcal{J}_{-j}\right) P^{\text{SEVI}}\left(Y = j | v\right) \\
&= E\left(\tilde{\varepsilon}_j 1\left\{v_k - v_j + \tilde{\varepsilon}_j \leq \tilde{\varepsilon}_k,\ k \in \mathcal{J}_{-j}\right\}\right) \\
&= E\left\{E\left(\tilde{\varepsilon}_j 1\left\{v_k - v_j + \tilde{\varepsilon}_j \leq \tilde{\varepsilon}_k,\ k \in \mathcal{J}_{-j}\right\} | \tilde{\varepsilon}_j\right)\right\} \\
&= \int_{-\infty}^{\infty} z \prod_{k \in \mathcal{J}_{-j}} \left[1 - G\left(v_k - v_j + z\right)\right] g\left(z\right) dz \\
&= \gamma + \sum_{\ell=1}^{J-1} \left(-1\right)^{\ell} \sum_{k_1 < \ldots < k_\ell : \{k_1, \ldots, k_\ell\} \subseteq \mathcal{J}_{-j}} \int_{-\infty}^{\infty} z G\left(v_{k_1} - v_j + z\right) \\
&\qquad \times G\left(v_{k_2} - v_j + z\right) \ldots G\left(v_{k_\ell} - v_j + z\right) g\left(z\right) dz,
\end{aligned}
$$

where $\gamma = \int_{-\infty}^{\infty} z g\left(z\right) dz$ is the Euler–Mascheroni constant.

For the integral in the preceding expression, we have

$$
\begin{aligned}
&\int_{-\infty}^{\infty} z G\left(v_{k_1} - v_j + z\right) G\left(v_{k_2} - v_j + z\right) \ldots G\left(v_{k_\ell} - v_j + z\right) g\left(z\right) dz \\
&= \int_{-\infty}^{\infty} z \exp\left[-\sum_{k \in \{k_1, \ldots, k_\ell\}} \exp\left(-v_k + v_j - z\right)\right] \exp\left[\left(-z - \exp\left(-z\right)\right)\right] dz \\
&= \int_{-\infty}^{\infty} z \exp\left[-\exp\left(-z\right) \sum_{k \in \{k_1, \ldots, k_\ell\}} s_{j|k}\right] \exp\left[\left(-z - \exp\left(-z\right)\right)\right] dz \\
&= -\int_{-\infty}^{\infty} z \exp\left[-\exp\left(-z\right) \sum_{k \in \{k_1, \ldots, k_\ell\}} s_{j|k}\right] \exp\left[\left(-\exp\left(-z\right)\right)\right] d\exp\left(-z\right).
\end{aligned}
$$

By employing a change of variable ($\tau = \exp(-z)$) and letting $\mathcal{S}_\ell := \{k_1, k_2, \ldots, k_\ell\}$, we then obtain

$$
\begin{aligned}
&\int_{-\infty}^{\infty} z G\left(v_{k_1} - v_j + z\right) G\left(v_{k_2} - v_j + z\right) \ldots G\left(v_{k_\ell} - v_j + z\right) g\left(z\right) dz \\
&= \int_0^{\infty} \ln\left(\frac{1}{\tau}\right) \exp\left[-\tau \sum_{k \in \mathcal{S}_\ell} s_{j|k}\right] \exp\left(-\tau\right) d\tau \\
&= \int_0^{\infty} \ln\left(\frac{1}{\tau}\right) \exp\left[-\tau \left(\sum_{k \in \mathcal{S}_\ell} s_{j|k} + 1\right)\right] d\tau \\
&= \frac{\gamma + \ln\left(\sum_{k \in \mathcal{S}_\ell} s_{j|k} + 1\right)}{\sum_{k \in \mathcal{S}_\ell} s_{j|k} + 1} = \frac{\gamma + v_j + \ln\left(\sum_{k \in \mathcal{S}_\ell} s_k^{-1} + s_j^{-1}\right)}{s_j \left(\sum_{k \in \mathcal{S}_\ell} s_k^{-1} + s_j^{-1}\right)} \\
&= \left[\gamma + v_j + \ln\left(\sum_{k \in \mathcal{S}_\ell \cup \{j\}} 1/s_k\right)\right] \frac{1/s_j}{\sum_{k \in \mathcal{S}_\ell \cup \{j\}} 1/s_k},
\end{aligned}
$$

where the second-to-last line follows because for any $\xi > 0$,

$$\begin{aligned}\int_0^\infty \ln(\frac{1}{\tau}) \exp(-\tau\xi)\, d\tau &= \frac{1}{\xi}\int_0^\infty \left[\ln(\frac{1}{\tau\xi}) + \ln(\xi)\right] \exp(-\tau\xi)\, d(\tau\xi) \\ &= \frac{\int_0^\infty \ln(\frac{1}{\tau}) \exp(-\tau)\, d\tau + \ln(\xi)}{\xi} = \frac{\gamma + \ln(\xi)}{\xi}.\end{aligned}$$

For notational simplicity, let $\mathcal{S}_{\ell,j} := \mathcal{S}_\ell \cup \{j\}$. Then

$$\begin{aligned}&E\left[\max_{j=1,\ldots,J}\{v_j + \varepsilon_j\}\right] \\ &= \sum_{j=1}^J v_j P^{\text{SEVI}}(Y = j|v) \\ &- \sum_{j=1}^J \left(\gamma + \sum_{\ell=1}^{J-1}(-1)^\ell \sum_{k_1<\ldots<k_\ell:\mathcal{S}_\ell=\{k_1,\ldots,k_\ell\}\subseteq\mathcal{J}_{-j}} \left[\gamma + v_j + \ln\left(\sum_{k\in\mathcal{S}_{\ell,j}} 1/s_k\right)\right] \frac{1/s_j}{\sum_{k\in\mathcal{S}_{\ell,j}} 1/s_k}\right) \\ &= \sum_{j=1}^J v_j P^{\text{SEVI}}(Y = j|v) - \gamma\sum_{j=1}^J \left[1 + \sum_{\ell=1}^{J-1}(-1)^\ell \sum_{k_1<\ldots<k_\ell:\mathcal{S}_\ell=\{k_1,\ldots,k_\ell\}\subseteq\mathcal{J}_{-j}} \frac{1/s_j}{\sum_{k\in\mathcal{S}_{\ell,j}} 1/s_k}\right] \\ &- \sum_{j=1}^J \sum_{\ell=1}^{J-1}(-1)^\ell \sum_{k_1<\ldots<k_\ell:\mathcal{S}_\ell=\{k_1,\ldots,k_\ell\}\subseteq\mathcal{J}_{-j}} \left[v_j + \ln\left(\sum_{k\in\mathcal{S}_{\ell,j}} 1/s_k\right)\right] \frac{1/s_j}{\sum_{k\in\mathcal{S}_{\ell,j}} 1/s_k} \\ &= \sum_{j=1}^J v_j \left[P^{\text{SEVI}}(Y = j|v) - \sum_{\ell=1}^{J-1}(-1)^\ell \sum_{k_1<\ldots<k_\ell:\mathcal{S}_\ell=\{k_1,\ldots,k_\ell\}\subseteq\mathcal{J}_{-j}} \frac{1/s_j}{\sum_{k\in\mathcal{S}_{\ell,j}} 1/s_k}\right] - \gamma \\ &- \sum_{j=1}^J \sum_{\ell=1}^{J-1}(-1)^\ell \sum_{k_1<\ldots<k_\ell:\mathcal{S}_\ell=\{k_1,\ldots,k_\ell\}\subseteq\mathcal{J}_{-j}} \frac{1/s_j}{\sum_{k\in\mathcal{S}_{\ell,j}} 1/s_k} \ln\left(\sum_{k\in\mathcal{S}_{\ell,j}} 1/s_k\right) \\ &= \left(\sum_{j=1}^J v_j - \gamma\right) - \sum_{j=1}^J \sum_{\ell=1}^{J-1}(-1)^\ell \sum_{k_1<\ldots<k_\ell:\mathcal{S}_\ell=\{k_1,\ldots,k_\ell\}\subseteq\mathcal{J}_{-j}} \frac{1/s_j}{\sum_{k\in\mathcal{S}_{\ell,j}} 1/s_k} \ln\left(\sum_{k\in\mathcal{S}_{\ell,j}} 1/s_k\right),\end{aligned}$$

where the third and last equalities use the result:

$$1 + \sum_{\ell=1}^{J-1}(-1)^\ell \sum_{k_1<\ldots<k_\ell:\mathcal{S}_\ell=\{k_1,\ldots,k_\ell\}\subseteq\mathcal{J}_{-j}} \frac{1/s_j}{\sum_{k\in\mathcal{S}_{\ell,j}} 1/s_k} = P^{\text{SEVI}}(Y = j|v).$$

Hence,

$$
\begin{aligned}
&E\left[\max_{j=1,\ldots,J}\{v_j+\varepsilon_j\}\right] \\
&= -\sum_{j=1}^{J}\sum_{\ell=0}^{J-1}(-1)^{\ell}\sum_{k_1<\ldots<k_\ell:\mathcal{S}_\ell=\{k_1,\ldots,k_\ell\}\subseteq\mathcal{J}_{-j}}\frac{1/s_j}{\sum_{k\in\mathcal{S}_{\ell,j}}1/s_k}\ln\left(\sum_{k\in\mathcal{S}_{\ell,j}}1/s_k\right)-\gamma \\
&= \sum_{\ell=1}^{J}(-1)^{\ell}\sum_{\mathcal{S}_\ell\subseteq\mathcal{J}}\left[\sum_{j\in\mathcal{S}_\ell}\frac{\exp(-v_j)}{\sum_{k\in\mathcal{S}_\ell}1/s_k}\right]\ln\left(\sum_{k\in\mathcal{S}_\ell}1/s_k\right)-\gamma \\
&= \sum_{\ell=1}^{J}(-1)^{\ell}\sum_{\mathcal{S}_\ell\subseteq\mathcal{J}}\ln\left(\sum_{k\in\mathcal{S}_\ell}1/s_k\right)-\gamma \\
&= \sum_{\ell=1}^{J}(-1)^{\ell}\sum_{|\mathcal{S}|=\ell,\mathcal{S}\subseteq\mathcal{J}}\ln\left[\sum_{k\in\mathcal{S}}1/s_k\right]-\gamma,
\end{aligned}
$$

where $\mathcal{S}_0=\varnothing$, the empty set, and the third line follows from exchanging the orders of summation and organizing the outer sum according to the value of $\ln\left(\sum_{k\in\mathcal{S}_{\ell,j}}1/s_k\right)$ for each $\mathcal{S}_\ell\subseteq\mathcal{J}_{-j}$. Using $\sum_{\ell=1}^{J}(-1)^{\ell}\sum_{|\mathcal{S}|=\ell,\mathcal{S}\subseteq\mathcal{J}}=-1$, we obtain the desired result.

The strict convexity of $W_J^{\mathrm{SEVI}}(v)$ follows from Theorem 6 of Sørensen and Fosgerau (2022), as it is easy to check that Conditions (C) and (S) in that paper hold. ■

S.B Proof of Theorem 3

Proof of Theorem 3. We prove Part (a) only. Part (b) follows from it, because $\widetilde{W}_N^{\mathrm{SEVI}}=\phi+\gamma$ by (11) and (12), and adding a constant affects neither Schur-convexity nor its strict form. Let $\mathcal{E}_1,\ldots,\mathcal{E}_N$ be iid standard exponential. The SEVI error satisfies $\varepsilon^{\mathrm{SEVI}}\overset{d}{=}\ln\mathcal{E}$, and the logarithm is increasing, so

$$
\phi(s)=W_N^{\mathrm{SEVI}}(\ln s)=E\max_j\left(\ln s_j+\varepsilon_j^{\mathrm{SEVI}}\right)=E\left[\ln\max_j s_j\mathcal{E}_j\right].
$$

Note that the set $\mathbb{R}_{++}^N$ is open, convex and invariant under permutations of the coordinates. Also, ϕ is unchanged by such a permutation and is smooth, being a finite sum of logarithms of sums of s_k^{-1} by Lemma 2. By the Schur–Ostrowski criterion for Schur-convexity (Marshall et al., 2011, Ch. 3, Thm. A.4), it therefore suffices to show that

$$
(s_i-s_j)\left(\frac{\partial\phi}{\partial s_i}-\frac{\partial\phi}{\partial s_j}\right)\geq 0 \text{ for all } i\neq j. \tag{S.3}
$$

For strict Schur-convexity, we use the second-order criterion of Marshall et al. (2011, Ch. 3, Thm. A.4.b): it suffices that, in addition to (S.3),

$$
\frac{\partial^2\phi}{\partial s_i^2}-2\frac{\partial^2\phi}{\partial s_i\partial s_j}+\frac{\partial^2\phi}{\partial s_j^2}>0 \tag{S.4}
$$

at every s with $\partial\phi/\partial s_i=\partial\phi/\partial s_j$. Steps 1–4 below establish (S.3), and Step 5 establishes (S.4) for $N\geq 3$.

Fix $i \neq j$, write $a = s_i$ and $b = s_j$, and assume $a > b$ without loss of generality, the case $a = b$ being trivial. Let

$$R(t) = \prod_{k \neq i,j} \left(1 - e^{-t/s_k}\right)$$

be the distribution function of $\max_{k \neq i,j} s_k \mathcal{E}_k$, with the convention $R \equiv 1$ when $N = 2$. R is nondecreasing, and strictly increasing on $(0, \infty)$ when $N \geq 3$. The distribution function of $\max_\ell s_\ell \mathcal{E}_\ell$ is then $F_s(t) = R(t)\left(1 - e^{-t/a}\right)\left(1 - e^{-t/b}\right)$.

Step 1: an integral representation and its derivative. For a positive random variable W with distribution function G_W satisfying $(\ln t)[1 - G_W(t)] \to 0$ as $t \to \infty$ and $(\ln t) G_W(t) \to 0$ as $t \downarrow 0$, integration by parts gives

$$E(\ln W) = \int_1^\infty \frac{1 - G_W(t)}{t} dt - \int_0^1 \frac{G_W(t)}{t} dt. \tag{S.5}$$

Both conditions hold for F_s. For the tail, $1 - F_s(t) = \Pr(\max_\ell s_\ell \mathcal{E}_\ell > t) \leq \sum_\ell e^{-t/s_\ell}$, so $(\ln t)[1 - F_s(t)] \to 0$ as $t \to \infty$. For the other condition, since $1 - e^{-x} \leq x$,

$$F_s(t) = \prod\nolimits_\ell \left(1 - e^{-t/s_\ell}\right) \leq \frac{t^N}{\prod_\ell s_\ell} \text{ for every } t > 0, \tag{S.6}$$

so that $F_s(t) = O\left(t^N\right)$, and hence $(\ln t) F_s(t) \to 0$, as $t \downarrow 0$. Applying (S.5) to $W = \max_\ell s_\ell \mathcal{E}_\ell$, we obtain

$$\phi(s) = \int_1^\infty \frac{1 - F_s(t)}{t} dt - \int_0^1 \frac{F_s(t)}{t} dt.$$

Using $\partial F_s(t)/\partial a = -\left(t/a^2\right) R(t)\left(1 - e^{-t/b}\right) e^{-t/a}$, we have

$$\frac{\partial \phi}{\partial a} = -\int_0^\infty \frac{1}{t} \frac{\partial F_s(t)}{\partial a} dt = \frac{1}{a^2} \int_0^\infty R(t)\left(1 - e^{-t/b}\right) e^{-t/a} dt, \tag{S.7}$$

where the interchange of derivative and integral is justified by dominated convergence. Similarly

$$\frac{\partial \phi}{\partial b} = -\int_0^\infty \frac{1}{t} \frac{\partial F_s(t)}{\partial b} dt = \frac{1}{b^2} \int_0^\infty R(t)\left(1 - e^{-t/a}\right) e^{-t/b} dt. \tag{S.8}$$

Step 2: a three-term kernel. Put $\alpha = 1/a$ and $\beta = 1/b$, so $0 < \alpha < \beta$. Taking the difference between (S.7) and (S.8) yields

$$\frac{\partial \phi}{\partial a} - \frac{\partial \phi}{\partial b} = \int_0^\infty R(t) k_{\alpha,\beta}(t) dt, \tag{S.9}$$

where

$$k_{\alpha,\beta}(t) = \left(\beta^2 - \alpha^2\right) e^{-(\alpha+\beta)t} - \beta^2 e^{-\beta t} + \alpha^2 e^{-\alpha t}. \tag{S.10}$$

Step 3: $k_{\alpha,\beta}$ crosses zero exactly once on $(0, \infty)$, from below. Integrating (S.10) term by term,

$$\int_0^\infty k_{\alpha,\beta}(t) dt = \frac{\beta^2 - \alpha^2}{\alpha + \beta} - \beta + \alpha = 0.$$

Also $k_{\alpha,\beta}(0) = \left(\beta^2 - \alpha^2\right) - \beta^2 + \alpha^2 = 0$, and

$$k'_{\alpha,\beta}(0) = -\left(\beta^2 - \alpha^2\right)(\alpha + \beta) + \beta^3 - \alpha^3 = \alpha\beta(\alpha - \beta) < 0.$$

Hence, $k_{\alpha,\beta} < 0$ on $(0, c)$ for some $c > 0$. On the other hand, since $\alpha < \beta < \alpha + \beta$, we have $e^{\alpha t} k_{\alpha,\beta}(t) \to \alpha^2 > 0$ as $t \to \infty$, so $k_{\alpha,\beta}(t) > 0$ for all sufficiently large t. Ordered by increasing exponent $-(\alpha+\beta) < -\beta < -\alpha$, the coefficients in (S.10) are $(\beta^2 - \alpha^2, -\beta^2, \alpha^2)$, all nonzero because $0 < \alpha < \beta$, with sign pattern $+, -, +$ and hence two sign changes. By the Descartes rule for exponential sums, a finite sum $\sum_r c_r e^{\lambda_r t}$ with strictly increasing exponents λ_r has at most as many real zeros, counted with multiplicity, as the coefficient sequence (c_r) has sign changes (Pólya and Szegő, 1976, Part Five, Ch. 1, Prob. 75), so $k_{\alpha,\beta}$ has at most two real zeros. The origin is one of them, and $k'_{\alpha,\beta}(0) \neq 0$ makes it simple, so at most one zero lies in $(0, \infty)$. By the intermediate value theorem, $k_{\alpha,\beta}$ has at least one zero in $(0, \infty)$. Hence, $k_{\alpha,\beta}$ has exactly one zero in $(0, \infty)$, say t^*, and $k_{\alpha,\beta} < 0$ on $(0, t^*)$ while $k_{\alpha,\beta} > 0$ on (t^*, ∞).

Step 4: signing the integral. In view of $\int_0^\infty k_{\alpha,\beta}(t)\, dt = 0$, we have

$$\frac{\partial \phi}{\partial a} - \frac{\partial \phi}{\partial b} = \int_0^\infty [R(t) - R(t^*)]\, k_{\alpha,\beta}(t)\, dt.$$

Since R is nondecreasing, $R(t) - R(t^*) \leq 0$ on $(0, t^*)$, where also $k_{\alpha,\beta}(t) \leq 0$, and $R(t) - R(t^*) \geq 0$ on (t^*, ∞), where also $k_{\alpha,\beta}(t) \geq 0$. The integrand is therefore nonnegative throughout, so the integral is nonnegative, and multiplying by $a - b > 0$ gives the Schur–Ostrowski inequality. Since i and j were arbitrary, ϕ is Schur-convex. For $N \geq 3$, the function R is strictly increasing, so $R(t) \neq R(t^*)$ for every $t \neq t^*$, while $k_{\alpha,\beta}(t) \neq 0$ for every $t \in (0, \infty)$ with $t \neq t^*$. The integrand is then strictly positive off t^*, so the integral is strictly positive and the quantity in (S.3) is strictly positive whenever $s_i \neq s_j$. For $N = 2$, we have $R \equiv 1$ and $\frac{\partial \phi}{\partial a} - \frac{\partial \phi}{\partial b}$ reduces to $\int_0^\infty k_{\alpha,\beta}(t)\, dt = 0$, so the Schur–Ostrowski quantity vanishes identically and ϕ is then both Schur-convex and Schur-concave. For $N = 1$, two vectors stand in the majorization order only if they are equal, so the claim holds trivially.

Step 5: the second-order condition. Let $N \geq 3$. By the strict inequality just established, $\partial\phi/\partial s_i = \partial\phi/\partial s_j$ forces $s_i = s_j$, so (S.4) need only be verified at $a = b$. The representations (S.7) and (S.8) hold for all $a, b > 0$, and the algebra leading to (S.10) does not use the ordering $a > b$, which enters only in Steps 3 and 4. At $a = b$, we have $\beta = \alpha$.

Next, we apply $\partial/\partial a - \partial/\partial b$ to (S.9). By the equality of mixed partial derivatives, the left-hand side becomes the quantity in (S.4), with $s_i = a$ and $s_j = b$, while on the right-hand side, R does not depend on a or b, so differentiating under the integral sign gives $\int_0^\infty R(t)\,(\partial/\partial a - \partial/\partial b)\, k_{\alpha,\beta}(t)\, dt$. This interchange is again justified by dominated convergence. Since $\alpha = 1/a$ and $\beta = 1/b$, we have $\partial/\partial a = -\alpha^2 \partial/\partial\alpha$ and $\partial/\partial b = -\beta^2 \partial/\partial\beta$. Differentiating $k_{\beta,\alpha} = -k_{\alpha,\beta}$ with respect to α shows that $\partial k_{\alpha,\beta}/\partial\beta = -\partial k_{\alpha,\beta}/\partial\alpha$ at $\beta = \alpha$, and a direct computation gives $\partial k_{\alpha,\beta}/\partial\alpha = -m_\alpha$, where

$$m_\alpha(t) = 2\alpha e^{-2\alpha t} - 2\alpha e^{-\alpha t} + \alpha^2 t e^{-\alpha t} = \alpha e^{-\alpha t}\left(2e^{-\alpha t} + \alpha t - 2\right).$$

Hence, $(\partial/\partial a - \partial/\partial b)\, k_{\alpha,\beta} = -2\alpha^2 \partial k_{\alpha,\beta}/\partial\alpha = 2\alpha^2 m_\alpha$ at $\beta = \alpha$, and therefore

$$\frac{\partial^2 \phi}{\partial a^2} - 2\frac{\partial^2 \phi}{\partial a \partial b} + \frac{\partial^2 \phi}{\partial b^2} = 2\alpha^2 \int_0^\infty R(t)\, m_\alpha(t)\, dt. \tag{S.11}$$

Integrating term by term yields $\int_0^\infty m_\alpha(t)\, dt = 1 - 2 + 1 = 0$. Writing $u = \alpha t$, m_α has the sign of $q(u) = 2e^{-u} + u - 2$. Here $q(0) = 0$, the derivative $q'(u) = 1 - 2e^{-u}$ is negative on $(0, \ln 2)$ and positive on $(\ln 2, \infty)$, and $q(u) \to \infty$, so q has exactly one zero u^* in $(0, \infty)$, and is negative on $(0, u^*)$ and positive on (u^*, ∞). Hence m_α crosses zero exactly once on $(0, \infty)$, from below, at $\tilde{t}^* = u^*/\alpha$. Subtracting $R(\tilde{t}^*) \int_0^\infty m_\alpha(t)\, dt = 0$ as in Step 4, the integral in (S.11) equals

$\int_0^\infty \left[R(t) - R\left(\tilde{t}^*\right)\right] m_\alpha(t)\, dt$. The integrand $\left[R(t) - R\left(\tilde{t}^*\right)\right] m_\alpha(t)$ is nonnegative throughout and, since R is strictly increasing when $N \geq 3$, it is positive off $\tilde{t}^*$. Therefore, (S.4) holds for the pair i, j. Since i and j were arbitrary, ϕ is strictly Schur-convex for $N \geq 3$. ■

S.C Proof of Proposition 1

The lemmas that follow concern the option-value gap $\Gamma_N(s)$ of (14), taken at a vector of $N \geq 1$ strictly positive scale values $s = (s_1, \ldots, s_N)$. By (11) together with Lemma 2,

$$\Gamma_N(s) = \ln\left(\sum\nolimits_{j=1}^{N} s_j\right) - \sum_{\ell=1}^{N} (-1)^\ell \sum_{|\mathcal{S}|=\ell, \mathcal{S}\subseteq\{1,\ldots,N\}} \ln\left[\sum_{j\in\mathcal{S}} s_j^{-1}\right], \tag{S.12}$$

and, by (12),

$$\Gamma_N(s) = \ln\left(\sum\nolimits_{j} s_j\right) - \phi(s) - \gamma. \tag{S.13}$$

As noted after Proposition 1, $\Gamma_N(s)$ is unchanged when every s_j is multiplied by a common constant.

The proof of Proposition 1(a) rests on the following four lemmas, which are also used in the proofs of Proposition 3(a) and (d). The first evaluates Γ_N for $N \leq 2$, the second combines with Theorem 3(a) to give the third, the nonnegativity of Γ_N, and the fourth upgrades this to strict positivity for $N \geq 3$.

Lemma S.1 *(i)* $\Gamma_1(s) = 0$ *for every* $s_1 > 0$*; (ii)* $\Gamma_2(s) = 0$ *for every pair* $s_1, s_2 > 0$.

Proof of Lemma S.1. For $N = 1$, (S.12) gives $\Gamma_1 = \ln s_1 - \left(-\ln s_1^{-1}\right) = 0$. For $N = 2$, the three nonempty subsets contribute $-\ln s_1^{-1}$, $-\ln s_2^{-1}$ and $+\ln\left(s_1^{-1} + s_2^{-1}\right)$, whose sum is

$$\ln s_1 + \ln s_2 + \ln \frac{s_1 + s_2}{s_1 s_2} = \ln(s_1 + s_2),$$

so the recentered SEVI surplus equals $\ln(s_1 + s_2)$, which is also the recentered LEVI surplus; hence $\Gamma_2 = 0$. ■

Lemma S.2 *Fix* $s = (s_1, \ldots, s_N)$ *with strictly positive entries. Then* $\Gamma_{N+1}(s, \varepsilon) \to \Gamma_N(s)$ *as* $\varepsilon \downarrow 0$. *More generally, if any subset of the coordinates tends to zero while the others remain fixed and positive,* Γ *converges to its value on the surviving coordinates. In particular, for* $\tau > 0$, $\Gamma_N(\tau, \varepsilon, \ldots, \varepsilon) \to \Gamma_1(\tau) = 0$.

Proof of Lemma S.2. Write $M = \max_{j\leq N} s_j \mathcal{E}_j$ and $M_\varepsilon = \max(M, \varepsilon\mathcal{E}_{N+1})$. By (S.13),

$$\Gamma_{N+1}(s, \varepsilon) = \ln\left(\varepsilon + \sum\nolimits_{j} s_j\right) - E[\ln M_\varepsilon] - \gamma,$$

and the first term converges to $\ln \sum_j s_j$ by continuity. Since $\varepsilon\mathcal{E}_{N+1} \to 0$ almost surely, $M_\varepsilon \to M$ almost surely. Fix $\varepsilon_0 > 0$. For every $\varepsilon \in (0, \varepsilon_0]$ we have $M \leq M_\varepsilon \leq \max(M, \varepsilon_0\mathcal{E}_{N+1})$, hence

$$|\ln M_\varepsilon| \leq |\ln M| + |\ln(\varepsilon_0\mathcal{E}_{N+1})|.$$

Both terms on the right are integrable: the second because $E\left|\ln \mathcal{E}\right| < \infty$, and the first because (S.6) bounds the distribution function of M by $t^N / \prod_j s_j$ near the origin while $1 - F_s$ decays exponentially at infinity. Dominated convergence gives $E\left[\ln M_\varepsilon\right] \to E\left[\ln M\right] = \phi\left(s\right)$, and the two limits combine to $\Gamma_N\left(s\right)$. The general statement is identical, with the surviving coordinates supplying the dominating bound. For the last claim, all coordinates but the first vanish, so $M_\varepsilon \to \tau\mathcal{E}_1$ almost surely, and $E\left[\ln\left(\tau\mathcal{E}_1\right)\right] = \ln\tau - \gamma$ gives $\Gamma_1\left(\tau\right) = \ln\tau - \left(\ln\tau - \gamma\right) - \gamma = 0$. ■

Theorem 3 is proved on the open set $\mathbb{R}^N_{++}$, so comparisons involving a boundary point are not immediately available. The next lemma supplies them, by approximating the boundary point from inside rather than by asserting that Schur-concavity extends.

Lemma S.3 $\Gamma_N\left(s\right) \geq 0$ *for every* $N \geq 1$ *and every* s *with strictly positive entries.*

Proof of Lemma S.3. Let $\tau = \sum_{j=1}^N s_j$ and for $\delta > 0$ set $s_\delta = \left(\tau - \left(N-1\right)\delta, \delta, \ldots, \delta\right)$, which has strictly positive entries for δ small and coordinate sum τ. Its sorted partial sums are $\tau - \left(N-k\right)\delta$ for $k = 1, \ldots, N$. Since s has strictly positive entries, $\sum_{i\leq k} s_{[i]} < \tau$ for every $k < N$, so for all sufficiently small δ we have $\tau - \left(N-k\right)\delta \geq \sum_{i\leq k} s_{[i]}$ for $k = 1, \ldots, N-1$, with equality of the full sums. Hence $s_\delta \succ s$. By (S.13) and Theorem 3(a), Γ_N is Schur-concave, so $\Gamma_N\left(s\right) \geq \Gamma_N\left(s_\delta\right)$. Write $c_\delta = \tau - \left(N-1\right)\delta$. Since Γ_N is unchanged when every coordinate is multiplied by a common constant,

$$\Gamma_N\left(s_\delta\right) = \Gamma_N\left(1, \delta/c_\delta, \ldots, \delta/c_\delta\right),$$

where the surviving coordinate is held at 1 while the remaining $N-1$ coordinates tend to zero, because $c_\delta \to \tau > 0$. Lemma S.2 therefore gives $\Gamma_N\left(s_\delta\right) \to \Gamma_1\left(1\right) = 0$ as $\delta \downarrow 0$. Hence $\Gamma_N\left(s\right) \geq 0$. ■

Lemma S.4 $\Gamma_N\left(s\right) > 0$ *for every* $N \geq 3$ *and every* $s = \left(s_1, \ldots, s_N\right)$ *with strictly positive entries.*

Proof of Lemma S.4. Order the coordinates so that $s_1 \geq \cdots \geq s_N$ and pick η with $0 < \eta < s_N$. Let s_η be the vector obtained from s by replacing s_1 with $s_1 + \eta$ and s_N with $s_N - \eta$. Then s_η has strictly positive entries and the same coordinate sum as s, and s is obtained from s_η by transferring η from the largest coordinate to the smallest, so $s_\eta \succ s$. Clearly, s_η is not a permutation of s. Since $N \geq 3$, (S.13) and Theorem 3(a) give strict Schur-concavity of Γ_N, so $\Gamma_N\left(s\right) > \Gamma_N\left(s_\eta\right)$. But Lemma S.3 gives $\Gamma_N\left(s_\eta\right) \geq 0$. Hence $\Gamma_N\left(s\right) > 0$. ■

Proof of Proposition 1. Part (a): Lemma S.1 gives $\Gamma_N\left(s\right) = 0$ for $N \leq 2$, and Lemma S.4 gives $\Gamma_N\left(s\right) > 0$ for $N \geq 3$.

Parts (b) and (c): since $\ln(\sum_j s_j)$ is unchanged when s is replaced by a vector with the same coordinate sum, (S.13) and Theorem 3(a) give that Γ_N is Schur-concave, strictly so for $N \geq 3$, which is Part (b). For Part (c), the equal vector is majorized by every vector with the same coordinate sum, and Γ_N is unchanged when every coordinate is multiplied by a common constant, so any s can be rescaled to that sum. ■

S.D Proofs of Propositions 2 and 3

Proof of Proposition 2. Part (a): Using $\widetilde{W}_J^{\text{LEVI}}(s) = \ln\left(\sum_{j=1}^J s_j\right)$, we have

$$\text{CV}^{\text{LEVI}}(1,s) = \ln\left(\sum_{j=1}^J s_j\right) - \ln\left(\sum_{j=2}^J s_j\right) = \ln\left(1 + \frac{s_1}{\sum_{j\in\mathcal{J}_{-1}} s_j}\right).$$

Part (b): By definition,

$$\widetilde{W}_J^{\text{SEVI}}(s) = \sum_{\ell=1}^J (-1)^\ell \sum_{|\mathcal{S}|=\ell, \mathcal{S}\subseteq\{1,\ldots,J\}} \ln\left[\sum_{j\in\mathcal{S}} s_j^{-1}\right],$$
$$\widetilde{W}_{J-1}^{\text{SEVI}}(s_{-1}) = \sum_{\ell=1}^{J-1} (-1)^\ell \sum_{|\mathcal{S}|=\ell, \mathcal{S}\subseteq\{2,\ldots,J\}} \ln\left[\sum_{j\in\mathcal{S}} s_j^{-1}\right].$$

Hence

$$\begin{aligned}
\text{CV}^{\text{SEVI}}(1,s) = \widetilde{W}_J^{\text{SEVI}}(s) - \widetilde{W}_{J-1}^{\text{SEVI}}(s_{-1}) &= \sum_{\ell=1}^J (-1)^\ell \sum_{|\mathcal{S}|=\ell, 1\in\mathcal{S}\subseteq\mathcal{J}} \ln\left[\sum_{j\in\mathcal{S}} s_j^{-1}\right] \\
&= \sum_{\ell=1}^J (-1)^\ell \sum_{|\mathcal{S}|=\ell, 1\in\mathcal{S}\subseteq\mathcal{J}} \ln\left[\left(s_1^{-1} + \sum_{j\in\mathcal{S}\setminus\{1\}} s_j^{-1}\right)\right] \\
&= \sum_{k=0}^{J-1} (-1)^{k+1} \sum_{|\mathcal{S}'|=k, \mathcal{S}'\subseteq\mathcal{J}_{-1}} \ln\left(1 + s_1 \sum_{j\in\mathcal{S}'} s_j^{-1}\right),
\end{aligned}$$

where the last line uses $\sum_{k=0}^{J-1}(-1)^{k+1}\sum_{|\mathcal{S}'|=k,\mathcal{S}'\subseteq\mathcal{J}_{-1}} 1 = \sum_{k=0}^{J-1}(-1)^{k+1}\binom{J-1}{k} = 0$. Noting that when $\mathcal{S}'$ is empty, $\sum_{j\in\mathcal{S}'} s_j^{-1} := 0$, we have

$$\text{CV}^{\text{SEVI}}(1,s) = \sum_{k=1}^{J-1} (-1)^{k+1} \sum_{|\mathcal{S}'|=k, \mathcal{S}'\subseteq\mathcal{J}_{-1}} \ln\left(1 + s_1 \sum_{j\in\mathcal{S}'} s_j^{-1}\right).$$

■

The proof of Proposition 3 requires one further lemma, which represents the compensating variation as an integral of choice probabilities along a "price" path, and underlies Part (b).

Lemma S.5 *Suppose $E\max_j|\varepsilon_j| < \infty$ and $\lambda = 1$. Then, for any distribution of the errors,*

$$\text{CV}(1,s) = E\left[\left(U_1 - \max_{j\neq 1} U_j\right)^+\right] = \int_0^\infty P\left(Y = 1 \mid v_1 - z, v_{-1}\right) dz, \tag{S.14}$$

where $x^+ = \max(x,0)$, $v_{-1} = (v_2,\ldots,v_J)$ and $s = \exp(v)$.

Proof of Lemma S.5. Write $M_{-1} = \max_{j \neq 1} U_j$. Removing alternative 1 lowers the expected maximum utility from $E \max(U_1, M_{-1})$ to EM_{-1}, so with $\lambda = 1$,

$$\mathrm{CV}(1, s) = E[\max(U_1, M_{-1}) - M_{-1}] = E\left[(U_1 - M_{-1})^+\right] = \int_0^\infty P(U_1 - M_{-1} > z)\, dz,$$

where the last step is the tail-integral formula for the expectation of a nonnegative random variable. For each $z > 0$, the event $\{U_1 - z > M_{-1}\}$ is exactly the event that alternative 1 is chosen when its systematic utility is lowered from v_1 to $v_1 - z$, the other systematic utilities being held fixed. Hence $P(U_1 - M_{-1} > z) = P(Y = 1 \mid v_1 - z, v_{-1})$, which gives (S.14). ■

Proof of Proposition 3. Part (a): Take $J = 3$ in (21), so that $D_{\mathrm{CV}}(1, s) = \Gamma_2(s_{-1}) - \Gamma_3(s)$. Lemma S.1(ii) gives $\Gamma_2(s_{-1}) = 0$, and Lemma S.4 with $N = 3$ gives $\Gamma_3(s) > 0$. Hence, $D_{\mathrm{CV}}(1, s) = -\Gamma_3(s) < 0$, that is, $\mathrm{CV}^{\mathrm{SEVI}}(1, s) < \mathrm{CV}^{\mathrm{LEVI}}(1, s)$.

Part (b): Since $s_j = \exp(v_j)$ and exp is increasing, the assumption $s_1 \leq \min_{j \neq 1} s_j$ is equivalent to $v_1 \leq \min_{j \neq 1} v_j$, and we argue in terms of the systematic utilities, to which Theorem 1 applies. Fix $z > 0$ and consider the systematic utility vector $(v_1 - z, v_2, \ldots, v_J)$. Since $v_1 - z < v_1 \leq \min_{j \neq 1} v_j$, alternative 1 is a strict laggard at that vector, so Theorem 1(b), which requires $J > 2$, applies and gives

$$P^{\mathrm{SEVI}}(Y = 1 \mid v_1 - z, v_{-1}) < P^{\mathrm{LEVI}}(Y = 1 \mid v_1 - z, v_{-1}) \qquad \text{for every } z > 0.$$

Note that the weak assumption $v_1 \leq \min_{j \neq 1} v_j$ suffices even though Theorem 1(b) requires a strict inequality, because lowering v_1 by any $z > 0$ moves the vector into the strict laggard cone immediately. The boundary configuration is a single point of the path and contributes nothing to the integral. Both sides are integrable over $z \in (0, \infty)$, since by Lemma S.5 the larger one integrates to $\mathrm{CV}^{\mathrm{LEVI}}(1, s) = \ln\left(1 + s_1 / \sum_{k \in \mathcal{J}_{-1}} s_k\right) < \infty$. Integrating the strict inequality over $z > 0$ and applying Lemma S.5 to each model gives $\mathrm{CV}^{\mathrm{SEVI}}(1, s) < \mathrm{CV}^{\mathrm{LEVI}}(1, s)$.

Part (c): For $\mathcal{S} \subseteq \mathcal{J}_{-1} = \{2, \ldots, J\}$, write $T_{\mathcal{S}} = \sum_{j \in \mathcal{S}} s_j^{-1}$. From the last display in Proposition 2(b), we have

$$\mathrm{CV}^{\mathrm{SEVI}}(1, s) = -\sum_{\mathcal{S} \subseteq \mathcal{J}_{-1}} (-1)^{|\mathcal{S}|} \ln(1 + s_1 T_{\mathcal{S}}) = -\sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} s_1^m A_m,$$

where $A_m := \sum_{\mathcal{S} \subseteq \mathcal{J}_{-1}} (-1)^{|\mathcal{S}|} T_{\mathcal{S}}^m$. Since

$$\sum_{\mathcal{S} \subseteq \mathcal{J}_{-1}} (-1)^{|\mathcal{S}|} e^{t T_{\mathcal{S}}} = \sum_{\mathcal{S} \subseteq \mathcal{J}_{-1}} (-1)^{|\mathcal{S}|} \sum_{m=0}^{\infty} \frac{t^m}{m!} T_{\mathcal{S}}^m = \sum_{m=0}^{\infty} \frac{t^m}{m!} A_m, \text{ and}$$

$$\sum_{\mathcal{S} \subseteq \mathcal{J}_{-1}} (-1)^{|\mathcal{S}|} e^{t T_{\mathcal{S}}} = \prod_{k=2}^{J} \left(1 - e^{t/s_k}\right) = (-1)^{J-1} t^{J-1} \prod_{k=2}^{J} s_k^{-1} + O\left(t^J\right),$$

as $t \to 0$, matching coefficients of $t^m/m!$ gives $A_m = 0$ for $m < J - 1$ and $A_{J-1} = (-1)^{J-1}(J - 1)! \prod_{k=2}^{J} s_k^{-1}$. Hence,

$$\mathrm{CV}^{\mathrm{SEVI}}(1, s) = -\frac{(-1)^J}{J - 1} s_1^{J-1} A_{J-1} + o\left(s_1^{J-1}\right) = (J - 2)! \left(\prod_{k=2}^{J} s_k^{-1}\right) s_1^{J-1} (1 + o(1)).$$

The rate for $\mathrm{CV}^{\mathrm{LEVI}}(1,s)$ follows immediately from Proposition 2(a), since $\ln(1+x) \sim x$ as $x \to 0$.

Part (d): Write $\Sigma_{-1} = \sum_{k\in\mathcal{J}_{-1}} s_k$, and recall $T_{\mathcal{S}} = \sum_{j\in\mathcal{S}} s_j^{-1}$ from the proof of Part (c). We proceed in five steps.

Step 1: the limit exists and equals $\Gamma_{J-1}(s_{-1})$. By Proposition 2(b),

$$\mathrm{CV}^{\mathrm{SEVI}}(1,s) = \sum_{\emptyset\neq\mathcal{S}\subseteq\mathcal{J}_{-1}} (-1)^{|\mathcal{S}|+1} \ln(1+s_1 T_{\mathcal{S}}),$$

and termwise $\ln(1+s_1T_{\mathcal{S}}) = \ln s_1 + \ln T_{\mathcal{S}} + O(s_1^{-1})$. Since $\sum_{\emptyset\neq\mathcal{S}\subseteq\mathcal{J}_{-1}}(-1)^{|\mathcal{S}|+1} = 1$, summing gives $\mathrm{CV}^{\mathrm{SEVI}}(1,s) = \ln s_1 + \sum_{\emptyset\neq\mathcal{S}}(-1)^{|\mathcal{S}|+1}\ln T_{\mathcal{S}} + O(s_1^{-1})$. By Proposition 2(a), $\mathrm{CV}^{\mathrm{LEVI}}(1,s) = \ln s_1 - \ln\Sigma_{-1} + O(s_1^{-1})$. Both diverge, but the $\ln s_1$ terms cancel in the difference, so the limit (20) exists and

$$D_{\mathrm{CV}}^{\infty} = \ln\Sigma_{-1} + \sum_{\emptyset\neq\mathcal{S}\subseteq\mathcal{J}_{-1}} (-1)^{|\mathcal{S}|+1}\ln T_{\mathcal{S}} = \Gamma_{J-1}(s_{-1}), \tag{S.15}$$

where the second equality follows from (S.12). So the asymptotic gap is exactly the option-value gap of the $J-1$ surviving alternatives.

Step 2: $J=2$ *and* $J=3$. These are $J-1=1$ and $J-1=2$ rivals, so $D_{\mathrm{CV}}^{\infty} = \Gamma_1 = 0$ and $D_{\mathrm{CV}}^{\infty} = \Gamma_2 = 0$ respectively, by Lemma S.1.

Step 3: $J \geq 4$. Apply Lemma S.4 with $N = J-1 \geq 3$ at the rival configuration s_{-1} to get $\Gamma_{J-1}(s_{-1}) > 0$. By Step 1, $D_{\mathrm{CV}}^{\infty} = \Gamma_{J-1}(s_{-1}) > 0$.

Step 4: Part (d)(ii). By (21) and (S.15),

$$D_{\mathrm{CV}}(1,s) = D_{\mathrm{CV}}^{\infty}(s_{-1}) - \Gamma_J(s).$$

Lemma S.3 gives $\Gamma_J(s) \geq 0$ for every $J \geq 2$, and Lemma S.4 with $N = J$ strengthens this to $\Gamma_J(s) > 0$ for $J \geq 3$. Hence $D_{\mathrm{CV}}(1,s) < D_{\mathrm{CV}}^{\infty}(s_{-1}) = \Gamma_{J-1}(s_{-1})$ for every $J \geq 3$.

Step 5: Part (d)(iii). Write $\Gamma_J(s) = \Gamma_J(s_1, s_{-1})$, regarded as a function of s_1 with s_{-1} held fixed, and consider (s_1, s_{-1}) and (s_1^+, s_{-1}) with $\max_{j\neq 1} s_j \leq s_1 < s_1^+$. Normalize each to unit coordinate sum. The first coordinate is the largest in both normalized vectors; passing from s_1 to s_1^+ raises it and scales every rival coordinate down by a common factor, so

$$\frac{1}{s_1^+ + \Sigma_{-1}}(s_1^+, s_{-1}) \succ \frac{1}{s_1 + \Sigma_{-1}}(s_1, s_{-1}),$$

and Proposition 1(b) gives

$$\Gamma_J\left[\frac{1}{s_1^+ + \Sigma_{-1}}(s_1^+, s_{-1})\right] \leq \Gamma_J\left[\frac{1}{s_1 + \Sigma_{-1}}(s_1, s_{-1})\right],$$

with strict inequality for $J \geq 3$, the two normalized vectors not being permutations of each other. Since Γ_J is unchanged when every coordinate is multiplied by a common positive constant, each side equals the corresponding unnormalized value, so $\Gamma_J(s_1^+, s_{-1}) \leq \Gamma_J(s_1, s_{-1})$, with strict inequality for $J \geq 3$. By (21), $D_{\mathrm{CV}}(1,s) = \Gamma_{J-1}(s_{-1}) - \Gamma_J(s)$ is therefore nondecreasing in s_1 on the leader region, and strictly increasing for $J \geq 3$. ∎

S.E Proofs of Theorems 4 and 5

The next lemma supplies the integrable envelope used in the proof of Theorem 4. It bounds the SEVI choice probability from below by a quantity that depends on the systematic utilities only through the largest advantage held over alternative j by any rival.

Lemma S.6 *Let $v \in \mathbb{R}^J$, let $j \in \mathcal{J}$, and set $\delta := \max_{k\in\mathcal{J}} (v_k - v_j)^+$. Then*

$$P^{\text{SEVI}}(Y = j|v) \geq \frac{(J-1)!}{J^J} e^{-(J-1)\delta}. \tag{S.16}$$

Consequently, for the conditional model $v_{ij}(\beta) = X_{ij}\beta$ and any compact $\mathcal{B} \subset \mathbb{R}^L$,

$$0 \leq \sup_{\beta\in\mathcal{B}} \ell_i^{\text{SEVI}}(\beta) \leq 2(J-1) \max_{k\in\mathcal{J}} \|X_{ik}\| \sup_{\beta\in\mathcal{B}} \|\beta\| + \log \frac{J^J}{(J-1)!}. \tag{S.17}$$

Proof of Lemma S.6. Write $r_k := s_{j|k} = \exp(v_j - v_k)$ for $k \in \mathcal{J}_{-j}$. Substituting $u = e^{-t}$ in (3) gives

$$P^{\text{SEVI}}(Y = j|v) = \int_0^\infty e^{-t} \prod_{k\in\mathcal{J}_{-j}} \left(1 - e^{-r_k t}\right) dt.$$

Define $\bar{r}_k := \min(r_k, 1)$ so that $\bar{r}_k \leq r_k$. Noting that $x \mapsto 1 - e^{-xt}$ is increasing for every $t > 0$, we have

$$\prod_{k\in\mathcal{J}_{-j}} \left(1 - e^{-r_k t}\right) \geq \prod_{k\in\mathcal{J}_{-j}} \left(1 - e^{-\bar{r}_k t}\right).$$

Using the elementary inequality $1 - e^{-x} \geq xe^{-x}$, which is $e^x - 1 \geq x$ after multiplying by e^x, we obtain

$$\prod_{k\in\mathcal{J}_{-j}} \left(1 - e^{-\bar{r}_k t}\right) \geq t^{J-1} \left(\prod_{k\in\mathcal{J}_{-j}} \bar{r}_k\right) \exp\left(-t \sum_{k\in\mathcal{J}_{-j}} \bar{r}_k\right).$$

Hence,

$$\begin{aligned} P^{\text{SEVI}}(Y = j|v) &\geq \left(\prod_{k\in\mathcal{J}_{-j}} \bar{r}_k\right) \int_0^\infty t^{J-1} \exp\left[-\left(1 + \sum_{k\in\mathcal{J}_{-j}} \bar{r}_k\right) t\right] dt \\ &= \left(\prod_{k\in\mathcal{J}_{-j}} \bar{r}_k\right) \frac{(J-1)!}{\left(1 + \sum_{k\in\mathcal{J}_{-j}} \bar{r}_k\right)^J}, \end{aligned}$$

where the last equality follows from the Gamma-function integral.

Since $\bar{r}_k \leq 1$ for all k, we have $\sum_{k\in\mathcal{J}_{-j}} \bar{r}_k \leq J - 1$ and $\left(1 + \sum_{k\in\mathcal{J}_{-j}} \bar{r}_k\right)^J \leq J^J$. For the product, we consider two cases for each fixed k. On the one hand, if $r_k \geq 1$, then $\bar{r}_k = 1 \geq e^{-\delta}$. On the other hand, if $r_k = \exp(v_j - v_k) < 1$, then $v_k - v_j > 0$, so that $\delta \geq v_k - v_j$ by the definition of δ in the statement. In this case, $\bar{r}_k = e^{-(v_k - v_j)} \geq e^{-\delta}$. Therefore, in both cases, $\bar{r}_k \geq e^{-\delta}$ for all k. Hence, $\prod_{k\in\mathcal{J}_{-j}} \bar{r}_k \geq e^{-(J-1)\delta}$. Putting all these together, we have

$$P^{\text{SEVI}}(Y = j|v) \geq \frac{(J-1)!}{J^J} e^{-(J-1)\delta},$$

completing the proof of (S.16).

For (S.17), the lower bound $\ell_i^{\mathrm{SEVI}}(\beta) = -\log P_{iY_i}^{\mathrm{SEVI}}(\beta) \geq 0$ holds because $P_{iY_i}^{\mathrm{SEVI}}(\beta) \leq 1$. Applying (S.16) at $v_i(\beta) = X_i\beta$ and $j = Y_i$ gives

$$\ell_i^{\mathrm{SEVI}}(\beta) \leq (J-1)\,\delta_i(\beta) + \log\frac{J^J}{(J-1)!}, \text{ where } \delta_i(\beta) = \max_{k\in\mathcal{J}}\left[(X_{ik} - X_{iY_i})\,\beta\right]^+.$$

By the Cauchy–Schwarz inequality and the triangle inequality,

$$\delta_i(\beta) \leq \left(\max_{k\in\mathcal{J}}\|X_{ik} - X_{iY_i}\|\right)\|\beta\| \leq 2\max_{k\in\mathcal{J}}\|X_{ik}\|\,\|\beta\|.$$

Taking the supremum over $\beta \in \mathcal{B}$ gives (S.17). ■

Proof of Theorem 4. Write $Q_n(\beta) := n^{-1}\ell_{\mathrm{SEVI}}(\beta)$ and $Q(\beta) := E\ell_i^{\mathrm{SEVI}}(\beta)$.

Step 1: uniform convergence. The random components have full support, so $P_{ik}^{\mathrm{SEVI}}(\beta) \in (0,1)$ for every β and every realization of X_i. Since P_{ik}^{SEVI} is continuous in β, ℓ_i^{SEVI} is also continuous in β. By Lemma S.6 and Assumption (ii) of the Theorem, the envelope (S.17) is integrable, its expectation being finite because $\sup_{\beta\in\mathcal{B}}\|\beta\| < \infty$ by the compactness in Assumption (i) of the Theorem. Under the iid sampling, we have the sufficient conditions for the uniform law of large numbers, and we have $\sup_{\beta\in\mathcal{B}}|Q_n(\beta) - Q(\beta)| \to^p 0$, and Q is finite and continuous on $\mathcal{B}$. See Lemma 2.4 of Newey and McFadden (1994) or Theorem 12.1 of Wooldridge (2002).

Step 2: β_0 uniquely minimizes Q. Conditioning on X_i and using $E[Y_{ik}|X_i] = P_{ik}^{\mathrm{SEVI}}(\beta_0)$, we have

$$Q(\beta) - Q(\beta_0) = E\left[\sum_{k=1}^{J} P_{ik}^{\mathrm{SEVI}}(\beta_0)\log\frac{P_{ik}^{\mathrm{SEVI}}(\beta_0)}{P_{ik}^{\mathrm{SEVI}}(\beta)}\right],$$

where both $Q(\beta)$ and $Q(\beta_0)$ are finite by Step 1. The inner sum is the Kullback–Leibler divergence between two distributions on $\mathcal{J}$, so it is nonnegative, and it vanishes if and only if $P_{ik}^{\mathrm{SEVI}}(\beta) = P_{ik}^{\mathrm{SEVI}}(\beta_0)$ for every k. Hence $Q(\beta) = Q(\beta_0)$ if and only if the two parameter values generate the same vector of choice probabilities almost surely.

By the identification result in Section 3.4, the map

$$v \mapsto \left[P^{\mathrm{SEVI}}(Y=1|v), \ldots, P^{\mathrm{SEVI}}(Y=J-1|v)\right]$$

is one-to-one on $\mathbb{R}_o^J$: the conjugate surplus function W_*^{SEVI} is well defined there, by the strict convexity of W_J^{SEVI} established in Lemma 2, and it recovers v from the choice probabilities through $v_j = \partial W_*^{\mathrm{SEVI}}(P)/\partial P_j$. Since choice probabilities depend on the systematic utilities only through their differences, equality of the two probability vectors is therefore equivalent to $V_i^*(\beta) = V_i^*(\beta_0)$ almost surely. Assumption (iii) of the Theorem, namely (23), rules this out for $\beta \neq \beta_0$. Hence $Q(\beta) > Q(\beta_0)$ for every $\beta \in \mathcal{B}$ with $\beta \neq \beta_0$.

Step 3: conclusion. $\mathcal{B}$ is compact by Assumption (i), $\beta_0 \in \mathcal{B}$, Q is continuous with a unique minimizer at β_0 by Step 2, and Q_n converges to Q uniformly on $\mathcal{B}$ by Step 1. Theorem 2.1 of Newey and McFadden (1994) then gives $\hat{\beta}_{\mathrm{MLE}} \to^p \beta_0$. ■

Proof of Theorem 5. Let $\mathbf{1}_J$ be the $J \times 1$ vector of ones. Since the choice probabilities do not change when a constant $\mu \in \mathbb{R}$ is added to each $V_{ij}(\beta)$, $E\sum_{k=1}^{J} Y_{ik}\log P^{\mathrm{Q}}(Y_i = k|\mathcal{J}, V_i(\beta) + \mu\mathbf{1}_J)$ is a constant function of μ. This implies that, as a function of μ, the derivative

of $E\sum_{k=1}^{J} Y_{ik} \log P^{\mathrm{Q}}(Y_i = k|\mathcal{J}, V_i(\beta) + \mu \mathbf{1}_J)$ with respect to μ is identically zero for any $V_i(\beta)$. But this derivative at $\mu = 0$ is

$$
\begin{aligned}
& E\sum_{k=1}^{J} Y_{ik} \frac{\partial \log P^{\mathrm{Q}}(Y_i = k|V_i(\beta) + \mu \mathbf{1}_J)}{\partial \mu}|_{\mu=0} \\
&= E\sum_{k=1}^{J} Y_{ik} \frac{\partial \log P^{\mathrm{Q}}(Y_i = k|V_i(\beta))}{\partial V_i(\beta)'} \mathbf{1}_J \\
&= E\left[\mathbf{1}_J' \sum_{k=1}^{J} Y_{ik} \frac{\partial \log P^{\mathrm{Q}}(Y_i = k|V_i(\beta))}{\partial V_i(\beta)}\right] = E\left[\mathbf{1}_J' \Lambda_i^{\mathrm{Q}}(X_i\beta) \overrightarrow{Y}_i\right].
\end{aligned}
$$

Hence, $E\left[\mathbf{1}_J' \Lambda_i^{\mathrm{Q}}(X_i\beta) \overrightarrow{Y}_i\right] = 0$ for any $\beta \in \mathbb{R}^L$. The perturbation above is in μ with β held fixed, so this is an identity in β rather than a first-order condition, and it requires no interiority. In particular, $E\left[\mathbf{1}_J' \Lambda_i^{\mathrm{Q}}\left(X_i\beta_{\mathrm{Q}}^*\right) \overrightarrow{Y}_i\right] = 0$. Combining this with Assumption (i) of the Theorem, we conclude that $\beta = \beta_{\mathrm{Q}}^*$ is the unique solution to

$$
\begin{aligned}
E\left[\mathbf{1}_J' \Lambda_i^{\mathrm{Q}}(X_i\beta) \overrightarrow{Y}_i\right] &= 0, \\
E\left[X_i' \Lambda_i^{\mathrm{Q}}(X_i\beta) \overrightarrow{Y}_i\right] &= 0.
\end{aligned}
$$

Since $\beta_0 \neq 0$, we know that $\beta_{0,\ell} \neq 0$ for some $\ell = 1, \ldots, L$. Without loss of generality, we assume that $\beta_{0,1} \neq 0$; this entails no loss because Assumption (iii) of the Theorem is imposed for every $\ell = 1, \ldots, L$ and is therefore invariant to a relabeling of the covariates. We partition X_i into $X_i = (X_i^{(1)}, X_i^{(2)}) \in \mathbb{R}^{J\times L}$ with $X_i^{(1)} \in \mathbb{R}^{J\times 1}$ and $X_i^{(2)} \in \mathbb{R}^{J\times(L-1)}$. For any $\beta \in \mathbb{R}^L$, partition it conformably as $\beta = (\beta_1, \beta_2')'$ where $\beta_1 \in \mathbb{R}$ and $\beta_2 \in \mathbb{R}^{L-1}$. In particular, we write $\beta_0 = \left(\beta_{0,1}, \beta_{0,2}'\right)'$ and $\beta_{\mathrm{Q}}^* = \left(\beta_{\mathrm{Q},1}^*, \beta_{\mathrm{Q},2}^{*\prime}\right)'$, where $\beta_{0,1}$ and $\beta_{\mathrm{Q},1}^*$ are the first elements of β_0 and β_{Q}^*, respectively. Define $\delta_1 = \beta_1/\beta_{0,1}$ and $\delta_2 = \beta_2 - \delta_1\beta_{0,2}$ so that there is a one-to-one and invertible map between β and $\delta = (\delta_1, \delta_2')'$. Then, we have the following reparametrization:

$$
\begin{aligned}
X_i\beta &= X_i\beta_0 + X_i(\beta - \beta_0) \\
&= X_i\beta_0 + X_i^{(1)}(\beta_1 - \beta_{0,1}) + X_i^{(2)}(\beta_2 - \beta_{0,2}) \\
&= X_i\beta_0 + X_i^{(1)}\beta_{0,1}(\delta_1 - 1) + X_i^{(2)}(\delta_2 + \delta_1\beta_{0,2} - \beta_{0,2}) \\
&= \left[X_i\beta_0 - X_i^{(1)}\beta_{0,1} - X_i^{(2)}\beta_{0,2}\right] + \left[X_i^{(1)}\beta_{0,1} + X_i^{(2)}\beta_{0,2}\right]\delta_1 + X_i^{(2)}\delta_2 \\
&= (X_i\beta_0)\,\delta_1 + X_i^{(2)}\delta_2 := \mathcal{X}_i\delta,
\end{aligned}
$$

where $\mathcal{X}_i = (\mathcal{X}_i^{(1)}, \mathcal{X}_i^{(2)})$, $\mathcal{X}_i^{(1)} = X_i\beta_0 \in \mathbb{R}^{J\times 1}$, and $\mathcal{X}_i^{(2)} = X_i^{(2)} \in \mathbb{R}^{J\times(L-1)}$. Note that

$$
\mathcal{X}_i = X_i \begin{pmatrix} \beta_{0,1} & \mathbf{0}_{1\times(L-1)} \\ \beta_{0,2} & I_{(L-1)\times(L-1)} \end{pmatrix}_{L\times L}.
$$

So, when $\beta_{0,1} \neq 0$, there is also a one-to-one and invertible map between $\mathcal{X}_i$ and X_i. As a result, the system of equations below

$$
\begin{pmatrix} E\left\{\left[X_i^{(1)}\right]' \Lambda_i^{\mathrm{Q}}(X_i\beta) \overrightarrow{Y}_i\right\} \\ E\left\{\left[X_i^{(2)}\right]' \Lambda_i^{\mathrm{Q}}(X_i\beta) \overrightarrow{Y}_i\right\} \end{pmatrix} = 0 \tag{S.18}
$$

is equivalent to

$$\begin{pmatrix} E\left\{\left[\mathcal{X}_i^{(1)}\right]' \Lambda_i^{\mathrm{Q}} \left(\mathcal{X}_i \delta\right) \overrightarrow{Y}_i\right\} \\ E\left\{\left[\mathcal{X}_i^{(2)}\right]' \Lambda_i^{\mathrm{Q}} \left(\mathcal{X}_i \delta\right) \overrightarrow{Y}_i\right\} \end{pmatrix} = 0. \tag{S.19}$$

Hence, by Assumption (i) of the Theorem, (S.19) also has a unique solution. Furthermore, if β_{Q}^* is the unique solution to (S.18), then $\delta^* = (\delta_1^*, \delta_2^{*\prime})'$ for

$$\delta_1^* = \frac{\beta_{\mathrm{Q},1}^*}{\beta_{0,1}},\ \delta_2^* = \beta_{\mathrm{Q},2}^* - \frac{\beta_{\mathrm{Q},1}^*}{\beta_{0,1}} \beta_{0,2},$$

is the unique solution to (S.19). Conversely, if $\delta^* = (\delta_1^*, \delta_2^{*\prime})'$ is the unique solution to (S.19), then $\beta_{\mathrm{Q}}^* = \left(\beta_{\mathrm{Q},1}^*, \beta_{\mathrm{Q},2}^{*\prime}\right)'$ for

$$\beta_{\mathrm{Q},1}^* = \delta_1^* \beta_{0,1},\ \beta_{\mathrm{Q},2}^* = \delta_2^* + \delta_1^* \beta_{0,2},$$

is the unique solution to (S.18).

We now show that (S.19) has a solution of the form $\delta_1 = \delta_1^\circ$ and $\delta_2 = \mathbf{0}$ for some scalar δ_1°. We choose $\delta_1 = \delta_1^\circ \neq 0$ such that the first equation in (S.19) holds when δ_2 is set equal to zero. That is, a nonzero δ_1° is chosen to satisfy

$$E\left\{\left[\mathcal{X}_i^{(1)}\right]' \Lambda_i^{\mathrm{Q}} (\mathcal{X}_i^{(1)} \delta_1^\circ) \overrightarrow{Y}_i\right\} = 0.$$

This is possible under Assumption (ii) of the Theorem. Note that

$$E\left[\overrightarrow{Y}_i | \mathcal{X}_i^{(1)}, \mathcal{X}_i^{(2)}\right] = E\left[\overrightarrow{Y}_i | X_i\beta_0, \mathcal{X}_i^{(2)}\right] = E\left[\overrightarrow{Y}_i | X_i\beta_0\right] = E\left[\overrightarrow{Y}_i | \mathcal{X}_i^{(1)}\right],$$

because $\overrightarrow{Y}_i$ is a function of $X_i\beta_0$ and ε_i only, and ε_i is independent of X_i. Now, by the law of iterated expectations, we have

$$\begin{aligned}
&E\left\{\left[\mathcal{X}_i^{(2)}\right]' \Lambda_i^{\mathrm{Q}} \left(\mathcal{X}_i^{(1)}\delta_1 + \mathcal{X}_i^{(2)}\delta_2\right) \overrightarrow{Y}_i\right\}_{\delta_1=\delta_1^\circ,\ \delta_2=\mathbf{0}} \\
&= E\left\{\left[\mathcal{X}_i^{(2)}\right]' \Lambda_i^{\mathrm{Q}} \left(\mathcal{X}_i^{(1)}\delta_1^\circ\right) \overrightarrow{Y}_i\right\} = E\left\{\left[\mathcal{X}_i^{(2)}\right]' \Lambda_i^{\mathrm{Q}} \left(\mathcal{X}_i^{(1)}\delta_1^\circ\right) E\left[\overrightarrow{Y}_i | \mathcal{X}_i^{(1)}, \mathcal{X}_i^{(2)}\right]\right\} \\
&= E\left\{\left[\mathcal{X}_i^{(2)}\right]' \Lambda_i^{\mathrm{Q}} \left(\mathcal{X}_i^{(1)}\delta_1^\circ\right) E\left[\overrightarrow{Y}_i | \mathcal{X}_i^{(1)}\right]\right\} = E\left\{E\left[\mathcal{X}_i^{(2)} | \mathcal{X}_i^{(1)}\right]' \Lambda_i^{\mathrm{Q}} \left(\mathcal{X}_i^{(1)}\delta_1^\circ\right) E\left[\overrightarrow{Y}_i | \mathcal{X}_i^{(1)}\right]\right\} \\
&= E\left\{E\left[\mathcal{X}_i^{(2)} | \mathcal{X}_i^{(1)}\right]' \Lambda_i^{\mathrm{Q}} \left(\mathcal{X}_i^{(1)}\delta_1^\circ\right) \overrightarrow{Y}_i\right\}.
\end{aligned}$$

Under Assumption (iii) of the Theorem, we have

$$\underbrace{E\left[\mathcal{X}_i^{(2)} | \mathcal{X}_i^{(1)}\right]}_{J\times(L-1)} = \underbrace{\mathcal{X}_i^{(1)}}_{J\times 1}\ \underbrace{\Theta_1'}_{1\times(L-1)} + \underbrace{\mathbf{1}_J}_{J\times 1}\ \underbrace{\Theta_0'}_{1\times(L-1)},$$

where $\Theta_0 = (\Theta_{0,2}, \ldots, \Theta_{0,L})'$ and $\Theta_1 = (\Theta_{1,2}, \ldots, \Theta_{1,L})'$. Using the above, $E\left\{\mathbf{1}_J' \Lambda_i^{\mathrm{Q}} \left(\mathcal{X}_i \delta\right) \overrightarrow{Y}_i\right\} = 0$ for any δ, and $E\left\{\left[\mathcal{X}_i^{(1)}\right]' \Lambda_i^{\mathrm{Q}} \left(\mathcal{X}_i^{(1)} \delta_1^\circ\right) \overrightarrow{Y}_i\right\} = 0$, we obtain

$$\begin{aligned}
&E\left\{\left[\mathcal{X}_i^{(2)}\right]' \Lambda_i^{\mathrm{Q}} \left(\mathcal{X}_i^{(1)} \delta_1 + \mathcal{X}_i^{(2)} \delta_2\right) \overrightarrow{Y}_i\right\}_{\delta_1 = \delta_1^\circ,\, \delta_2 = \mathbf{0}} \\
&= \Theta_0 E\left\{\mathbf{1}_J' \Lambda_i^{\mathrm{Q}} \left(\mathcal{X}_i^{(1)} \delta_1^\circ\right) \overrightarrow{Y}_i\right\} + \Theta_1 E\left\{\left[\mathcal{X}_i^{(1)}\right]' \Lambda_i^{\mathrm{Q}} \left(\mathcal{X}_i^{(1)} \delta_1^\circ\right) \overrightarrow{Y}_i\right\} = 0.
\end{aligned}$$

We have therefore shown that $\delta = \left(\delta_1^\circ, \mathbf{0}_{1\times(L-1)}\right)'$ is a solution to (S.19).

Since (S.19) has a unique solution, we must have

$$\delta_1^* = \delta_1^\circ,\ \delta_2^* = \mathbf{0}.$$

Therefore, the unique solution β_{Q}^* satisfies

$$\beta_{\mathrm{Q},1}^* = \delta_1^* \beta_{0,1},\ \beta_{\mathrm{Q},2}^* = \delta_1^* \beta_{0,2},$$

that is, $\beta_{\mathrm{Q}}^* = \delta_1^* \beta_0 = \delta_1^\circ \beta_0$, as desired. ■

S.F Algebraic Relation between SEVI and LEVI Choice Probabilities

To relate the choice probabilities under the SEVI and LEVI assumptions, we consider a general latent utility model $U_{ij} = V_{ij} + \varepsilon_{ij}$, $j = 1, \ldots, J$ where ε_{ij} is iid across $j = 1, \ldots, J$ and may follow either the SEVI or LEVI distribution. We assume that $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iJ})'$ is independent of $V_i = (V_{i1}, \ldots, V_{iJ})'$. For any alternative j and any subset of alternatives $\mathcal{S}_\ell$ that does not include alternative j, we let

$$\begin{aligned}
P^{\mathrm{SEVI}}(Y_i = j | \mathcal{S}_\ell \cup \{j\}, V_i = v_i) &= \Pr(Y_i = j | \mathcal{S}_\ell \cup \{j\}, V_i = v_i, \varepsilon_{im} \sim \textit{iid SEVI over } m \in \mathcal{J}), \\
P^{\mathrm{LEVI}}(Y_i = j | \mathcal{S}_\ell \cup \{j\}, V_i = v_i) &= \Pr(Y_i = j | \mathcal{S}_\ell \cup \{j\}, V_i = v_i, \varepsilon_{im} \sim \textit{iid LEVI over } m \in \mathcal{J}),
\end{aligned}$$

which are the probabilities of individual i choosing alternative j from the subset $\mathcal{S}_\ell \cup \{j\}$ under SEVI and LEVI specifications, respectively. In this definition, the probabilities depend on the specific choice subset under consideration.

Proposition S.1 *The choice probabilities under the SEVI and LEVI errors satisfy the following relations:*

$$\begin{aligned}
&P^{\mathrm{SEVI}}(Y_i = j | \mathcal{J}, V_i = v_i) \\
&\qquad = 1 + \sum_{\ell=1}^{J-1} (-1)^\ell \sum_{k_1 < \ldots < k_\ell:\ \mathcal{S}_\ell = \{k_1, \ldots, k_\ell\} \subseteq \mathcal{J}_{-j}} P^{\mathrm{LEVI}}\left(Y_i = j | \mathcal{S}_\ell \cup \{j\}, V_i = -v_i\right);
\end{aligned}$$

$$\begin{aligned}
&P^{\mathrm{LEVI}}(Y_i = j | \mathcal{J}, V_i = v_i) \\
&\qquad = 1 + \sum_{\ell=1}^{J-1} (-1)^\ell \sum_{k_1 < \ldots < k_\ell:\ \mathcal{S}_\ell = \{k_1, \ldots, k_\ell\} \subseteq \mathcal{J}_{-j}} P^{\mathrm{SEVI}}\left(Y_i = j | \mathcal{S}_\ell \cup \{j\}, V_i = -v_i\right).
\end{aligned}$$

Proof of Proposition S.1. The first result follows from Lemma 1:

$$\begin{aligned}
P^{\text{SEVI}}\left(Y_i=j|v_i\right) &= 1+\sum_{\ell=1}^{J-1}\left(-1\right)^{\ell}\sum_{k_1<...<k_\ell:\{k_1,...,k_\ell\}\subseteq\mathcal{J}_{-j}}\frac{1}{1+\sum_{k\in\{k_1,...,k_\ell\}}\exp\left(v_{ij}-v_{ik}\right)}\\
&= 1+\sum_{\ell=1}^{J-1}\left(-1\right)^{\ell}\sum_{k_1<...<k_\ell:\{k_1,...,k_\ell\}\subseteq\mathcal{J}_{-j}}\frac{\exp\left(-v_{ij}\right)}{\exp\left(-v_{ij}\right)+\sum_{k\in\{k_1,...,k_\ell\}}\exp\left(-v_{ik}\right)}\\
&= 1+\sum_{\ell=1}^{J-1}\left(-1\right)^{\ell}\sum_{k_1<...<k_\ell:\mathcal{S}_\ell:=\{k_1,...,k_\ell\}\subseteq\mathcal{J}_{-j}}P^{\text{LEVI}}\left(Y_i=j|\mathcal{S}_\ell\cup\{j\},V_i=-v_i\right).
\end{aligned}$$

To prove the second result, we note that

$$\begin{aligned}
&P^{\text{LEVI}}\left(Y_i=j|\mathcal{J},V_i=v_i\right)\\
&=\Pr\left(\cap_{k\in\mathcal{J}_{-j}}\left\{V_{ij}+\varepsilon_{ij}\geq V_{ik}+\varepsilon_{ik}\right\}|V_i=v_i;\varepsilon_{im}\sim iid\text{ LEVI for }m\in\mathcal{J}\right)\\
&=\Pr\left(\cap_{k\in\mathcal{J}_{-j}}\left\{-V_{ij}-\varepsilon_{ij}\leq -V_{ik}-\varepsilon_{ik}\right\}|V_i=v_i;\varepsilon_{im}\sim iid\text{ LEVI for }m\in\mathcal{J}\right)\\
&=\Pr\left(\cap_{k\in\mathcal{J}_{-j}}\left\{V_{ij}+\varepsilon_{ij}\leq V_{ik}+\varepsilon_{ik}\right\}|V_i=-v_i;\varepsilon_{im}\sim iid\text{ SEVI for }m\in\mathcal{J}\right)\\
&=1-\Pr\left(\left[\cap_{k\in\mathcal{J}_{-j}}\left\{V_{ij}+\varepsilon_{ij}\leq V_{ik}+\varepsilon_{ik}\right\}\right]^c|V_i=-v_i;\varepsilon_{im}\sim iid\text{ SEVI for }m\in\mathcal{J}\right)\\
&=1-\Pr\left(\cup_{k\in\mathcal{J}_{-j}}\left[\left\{V_{ij}+\varepsilon_{ij}\leq V_{ik}+\varepsilon_{ik}\right\}^c\right]|V_i=-v_i;\varepsilon_{im}\sim iid\text{ SEVI for }m\in\mathcal{J}\right)\\
&=1-\Pr\left(\cup_{k\in\mathcal{J}_{-j}}\left[\left\{V_{ij}+\varepsilon_{ij}>V_{ik}+\varepsilon_{ik}\right\}\right]|V_i=-v_i;\varepsilon_{im}\sim iid\text{ SEVI for }m\in\mathcal{J}\right)\\
&=1-\Pr\left(\cup_{k\in\mathcal{J}_{-j}}\mathfrak{E}_{ik}|V_i=-v_i;\varepsilon_{im}\sim iid\text{ SEVI for }m\in\mathcal{J}\right),
\end{aligned}$$

where $\mathfrak{E}_{ik}$ is the event $\{V_{ij}+\varepsilon_{ij}>V_{ik}+\varepsilon_{ik}\}$. Using the Inclusion-Exclusion Principle, which states that

$$\Pr\left(\cup_{k\in\mathcal{J}_{-j}}\mathfrak{E}_{ik}|V_i=-v_i\right)=\sum_{\ell=1}^{J-1}\left(-1\right)^{\ell-1}\sum_{k_1<...<k_\ell:\mathcal{S}_\ell=\{k_1,...,k_\ell\}\subseteq\mathcal{J}_{-j}}\Pr\left(\cap_{k\in\mathcal{S}_\ell}\mathfrak{E}_{ik}|V_i=-v_i\right),$$

we obtain

$$\begin{aligned}
&P^{\text{LEVI}}\left(Y_i=j|\mathcal{J},V_i=v_i\right)\\
&=1+\sum_{\ell=1}^{J-1}\left(-1\right)^{\ell}\sum_{k_1<...<k_\ell:\mathcal{S}_\ell=\{k_1,...,k_\ell\}\subseteq\mathcal{J}_{-j}}P^{\text{SEVI}}\left(Y_i=j|\mathcal{S}_\ell\cup\{j\},V_i=-v_i\right).
\end{aligned}$$

■

This proposition reveals an intrinsic symmetry: the SEVI probability for a set of systematic utilities v_i is a signed inclusion-exclusion sum of LEVI probabilities under the reflected utilities $-v_i$, and vice versa. This symmetry is not tied specifically to Type I extreme-value distributions; it is a general property for any pair of distributions that are mirror images of each other. In fact, such a property holds even beyond RUMs; see Theorem 1 of de Palma and Kilani (2023).

Taking the difference between the two equations in Proposition S.1 yields:

$$\begin{aligned}
&P^{\text{SEVI}}(Y_i=j|\mathcal{J},V_i=v_i)-P^{\text{LEVI}}(Y_i=j|\mathcal{J},V_i=v_i)\\
&\qquad=\sum_{\ell=2}^{J-1}\left(-1\right)^{\ell-1}\sum_{k_1<...<k_\ell:\ \mathcal{S}_\ell=\{k_1,...,k_\ell\}\subseteq\mathcal{J}_{-j}}\left[P^{\Delta}\left(Y_i=j|\mathcal{S}_\ell\cup\{j\},V_i=-v_i\right)\right],
\end{aligned}$$

where

$$P^{\Delta}\left(Y_i = j|\mathcal{S}_\ell \cup \{j\}, V_i = -v_i\right)$$
$$= P^{\text{SEVI}}\left(Y_i = j|\mathcal{S}_\ell \cup \{j\}, V_i = -v_i\right) - P^{\text{LEVI}}\left(Y_i = j|\mathcal{S}_\ell \cup \{j\}, V_i = -v_i\right).$$

Thus, the difference in the SEVI and LEVI probabilities of choosing alternative j from the entire choice set $\mathcal{J}$ is an aggregation of the probability differences from choosing alternative j from all subsets of alternatives that contain j, but with flipped systematic utilities. This is a recursive relationship, as each probability difference on the right-hand side can be represented in the same fashion as that on the left-hand side.

Replacing v_i with $-v_i$ in the second relation of Proposition S.1 yields:

$$P^{\text{LEVI}}(Y_i = j|\mathcal{J}, V_i = -v_i)$$
$$= 1 + \sum_{\ell=1}^{J-1} (-1)^\ell \sum_{k_1<\ldots<k_\ell:\mathcal{S}_\ell:=\{k_1,\ldots,k_\ell\}\subseteq\mathcal{J}_{-j}} P^{\text{SEVI}}\left(Y_i = j|\mathcal{S}_\ell \cup \{j\}, V_i = v_i\right). \quad \text{(S.20)}$$

The right-hand side is a special case of a Block–Marschak polynomial. In stochastic choice theory, the nonnegativity of these polynomials is both necessary and sufficient for a choice rule to have an additive random utility representation (Strzalecki, 2023). While the SEVI model satisfies this trivially as an additive RUM, equation (S.20) is noteworthy because it provides a simple characterization of this polynomial through the LEVI–SEVI duality.

More generally, let j be a target alternative of interest and

$$\mathcal{D} := \{d_1, \ldots, d_\rho\} \subseteq \mathcal{J}_{-j} \text{ for some } \rho \leq J-1$$

be a subset of alternatives that excludes alternative j. Then, the general Block–Marschak polynomial is defined as:

$$\mathcal{K}^{\text{M}}(j, \mathcal{D}|v_i)$$
$$= P^{\text{M}}\left(Y_i = j|\mathcal{D} \cup \{j\}, V_i = v_i\right)$$
$$+ \sum_{\ell=1}^{J-\rho-1} (-1)^\ell \sum_{k_1<\ldots<k_\ell:\mathcal{S}_\ell:=\{k_1,\ldots,k_\ell\}\subseteq\mathcal{J}_{-j}\setminus\mathcal{D}} P^{\text{M}}\left(Y_i = j|\mathcal{S}_\ell \cup \mathcal{D} \cup \{j\}, V_i = v_i\right)$$

for $\text{M} \in \{\text{SEVI}, \text{LEVI}\}$. As noted by Barberá and Pattanaik (1986), $\mathcal{K}^{\text{M}}(j, \mathcal{D}|v_i)$ represents the probability that alternative j dominates every alternative in $\mathcal{D}$ but is dominated by every alternative in the complement $\mathcal{J}_{-j}\setminus\mathcal{D}$. Hence, $\mathcal{K}^{\text{M}}(j, \mathcal{D}|v_i) \geq 0$ for all $(j, \mathcal{D})$, consistent with the nonnegativity of Block–Marschak polynomials for any additive RUM. The quantity $\mathcal{K}^{\text{M}}(j, \mathcal{D}|v_i)$ is also the key probability for rank-ordered data: the probability that alternative j is ranked above every alternative in $\mathcal{D}$ but below all the others.

The duality established in Proposition S.1 extends to the entire space of Block–Marschak polynomials. The probability that j dominates $\mathcal{D}$ but is dominated by $\mathcal{J}_{-j}\setminus\mathcal{D}$ under systematic utilities v_i in a SEVI model is the same as the probability that j is dominated by $\mathcal{D}$ but dominates $\mathcal{J}_{-j}\setminus\mathcal{D}$ under a LEVI model with flipped utilities $-v_i$. Consequently, we have

$$\mathcal{K}^{\text{SEVI}}(j, \mathcal{D}|v_i) = \mathcal{K}^{\text{LEVI}}(j, \mathcal{J}_{-j}\setminus\mathcal{D}| - v_i),$$
$$\mathcal{K}^{\text{LEVI}}(j, \mathcal{D}|v_i) = \mathcal{K}^{\text{SEVI}}(j, \mathcal{J}_{-j}\setminus\mathcal{D}| - v_i).$$

Taking $\mathcal{D}$ to be the empty set or $\mathcal{J}_{-j}$ yields the result in Proposition S.1.

This dual representation shows that for rank-ordered data, both LEVI and SEVI specifications involve computing sums of logit-type terms. Consequently, the computational advantage of the standard LEVI logit model compared to the SEVI model does not carry over to rank-ordered data. From a computational standpoint, the two specifications are comparable, and the choice between them should be based on which model better characterizes reality, rather than on computational convenience. While extending the SEVI specification to rank-ordered data remains an area for future research, the above duality results provide the analytical foundation.

S.G Further Comparisons of the SEVI and LEVI Models

This subsection collects two comparisons that sharpen and explain the "Leader-Takes-More" effect of Theorem 1. The first specializes to $J = 3$, where the difference between the two choice probabilities is available in closed form, so that both its sign and its magnitude can be characterized exactly. The second traces the effect to the largest advantage that any rival draws over a fixed alternative, the single random quantity that settles whether that alternative is chosen.

S.G.1 The probability difference when $J = 3$: sign, magnitude, and a market illustration

Theorem 1 requires alternative 1 to be ranked in the same direction against every competitor. For $J = 3$ the difference between the two choice probabilities can be evaluated in closed form, and this delivers a complete characterization, under which neither regime need hold.

Proposition S.2 *Let $J = 3$. Then $P^{\mathrm{SEVI}}(Y = 1|v) > P^{\mathrm{LEVI}}(Y = 1|v)$ if and only if $v_1 > (v_2 + v_3)/2$, and $P^{\mathrm{SEVI}}(Y = 1|v) = P^{\mathrm{LEVI}}(Y = 1|v)$ if and only if $v_1 = (v_2 + v_3)/2$.*

Proof of Proposition S.2. Let $s_2 = \exp(v_2 - v_1)$ and $s_3 = \exp(v_3 - v_1)$. By definition,

$$P^{\mathrm{SEVI}}(Y = 1|v) = 1 - \frac{1}{1 + 1/s_2} - \frac{1}{1 + 1/s_3} + \frac{1}{1 + 1/s_2 + 1/s_3},$$
$$P^{\mathrm{LEVI}}(Y = 1|v) = \frac{1}{1 + s_2 + s_3}.$$

The difference is given by

$$P^{\mathrm{SEVI}}(Y = 1|v) - P^{\mathrm{LEVI}}(Y = 1|v) = \frac{s_2 s_3 (1 - s_2 s_3)}{(s_2 + 1)(s_3 + 1)(s_2 + s_3 + 1)(s_2 + s_3 + s_2 s_3)}.$$

Every factor other than $1 - s_2 s_3$ is strictly positive, so the sign of the difference is that of $1 - s_2 s_3$ alone. Since $s_2 s_3 = \exp(v_2 + v_3 - 2v_1)$, we have $1 - s_2 s_3 > 0$ if and only if $v_1 > (v_2 + v_3)/2$, and $1 - s_2 s_3 = 0$ if and only if $v_1 = (v_2 + v_3)/2$. ■

Proposition S.2 is sharper than Theorem 1 at $J = 3$ in two respects: it is an equivalence rather than a one-way implication, and it compares v_1 with the average of the two competing utilities rather than with each of them separately, so that a competitor stronger than alternative 1 is permitted provided the other is correspondingly weaker. No comparable closed form is available for $J > 3$, where the sign of the difference is not governed by $\prod_{j \geq 2} s_j$ alone.

For $J = 3$ itself, the closed form obtained in the proof of Proposition S.2 delivers the magnitude of the difference as well as its sign. To display it, define the *scale values* $s = (s_1, s_2, s_3)'$ with $s_j = \exp(v_j)$ and fix $s_1 = 1$, so that the s_2 and s_3 of that proof are the scale values of the second and third alternatives and

$$P^{\text{SEVI}}(Y = 1|v) - P^{\text{LEVI}}(Y = 1|v) = \frac{s_2 s_3 (1 - s_2 s_3)}{(s_2 + 1)(s_3 + 1)(s_2 + s_3 + 1)(s_2 + s_3 + s_2 s_3)}.$$

We create a contour plot of this difference against (s_2, s_3). We also make a plot of the probability difference against s_2 while holding $s_2 + s_3$ constant at a level C; that is, setting $s_3 = C - s_2$. Figures S.1 and S.2 present these two plots. Both figures confirm that the difference is a highly nonlinear function of s_2 and s_3. Its sign, however, is governed by the product alone: Proposition S.2 with $s_1 = 1$ gives $P^{\text{SEVI}}(Y = 1|v) > P^{\text{LEVI}}(Y = 1|v)$ if and only if $s_2 s_3 < 1$, so the zero contour in Figure S.1 is the hyperbola $s_2 s_3 = 1$.

It is instructive to compare the criterion $s_2 s_3 < 1$ with another criterion that a logit user may use. Since $P^{\text{LEVI}}(Y = 1|v) = (1 + s_2 + s_3)^{-1}$, alternative 1 commands more than an equal share of the market under the LEVI, that is, $P^{\text{LEVI}}(Y = 1|v) > 1/3$, exactly when $(s_2 + s_3)/2 < 1$. Since the sum is the only feature of the rivals that the logit ever sees, this makes $(s_2 + s_3)/2 < 1$ the natural reference point. Both conditions are in fact comparisons of an average of the rivals' scale values with $s_1 = 1$: the logit share criterion $(s_2 + s_3)/2 < 1$ places their arithmetic mean below s_1, while the criterion of Proposition S.2, $\sqrt{s_2 s_3} < 1$, places their geometric mean below s_1.

Because the geometric mean never exceeds the arithmetic mean, $\sqrt{s_2 s_3} \leq (s_2 + s_3)/2$, the logit share criterion implies that of Proposition S.2: $(s_2 + s_3)/2 < 1$ forces $\sqrt{s_2 s_3} < 1$. The converse fails, and it fails exactly where the two means separate, that is, where the rivals are unequal. At $(s_2, s_3) = (4, 0.2)$, for instance, the arithmetic mean is 2.1 while the geometric mean is 0.894. Alternative 1 is a distant second under either model, holding $P^{\text{LEVI}}(Y = 1|v) = 0.192$ against alternative 2's 0.769, and yet the SEVI still raises it, to $P^{\text{SEVI}}(Y = 1|v) = 0.193$ against alternative 2's 0.792. An alternative can therefore gain from the SEVI while trailing badly, so long as its rivals are lopsided enough that their geometric mean stays below its own scale value.

The magnitude of the difference, unlike its sign, depends on the sum as well. For a fixed $C = s_2 + s_3$, it depends on (s_2, s_3) only through the product $p = s_2 s_3$, because $(s_2 + 1)(s_3 + 1) = p + C + 1$; that is,

$$P^{\text{SEVI}}(Y = 1|v) - P^{\text{LEVI}}(Y = 1|v) = \frac{p(1 - p)}{(C + 1)(p + C + 1)(C + p)}. \tag{S.21}$$

For a fixed sum C, the product p is largest when the two rivals are evenly matched, at $s_2 = s_3 = C/2$, and (S.21) is increasing in p over the feasible range $0 < p \leq C^2/4$ precisely when $C < \sqrt{5} - 1 \approx 1.236$.[14] For such C, the closer s_2 and s_3 are to one another, the larger the difference $P^{\text{SEVI}}(Y = 1|v) - P^{\text{LEVI}}(Y = 1|v)$ becomes: with the combined strength of the rivals held fixed below $\sqrt{5} - 1$, the SEVI favors alternative 1 most when neither rival stands out. In summary, in line with Figures S.4–S.6 of Section S.H, Figures S.1 and S.2 show that the SEVI model produces a higher choice probability for the most favored alternative but a lower choice probability for the least favored alternative than the LEVI model does.

[14] Differentiating (S.21) with respect to p shows that the sign of the derivative is that of $C - 2Cp - 2p^2$, so the difference is increasing in p if and only if $2p^2 + 2Cp - C < 0$. Because the difference is increasing and then decreasing in p, it is increasing throughout $0 < p \leq C^2/4$ if and only if this holds at $p = C^2/4$, where the left-hand side equals $C(C + 2)(C^2 + 2C - 4)/8$. As $C > 0$, the condition reduces to $C^2 + 2C - 4 < 0$, that is, $C < \sqrt{5} - 1$.

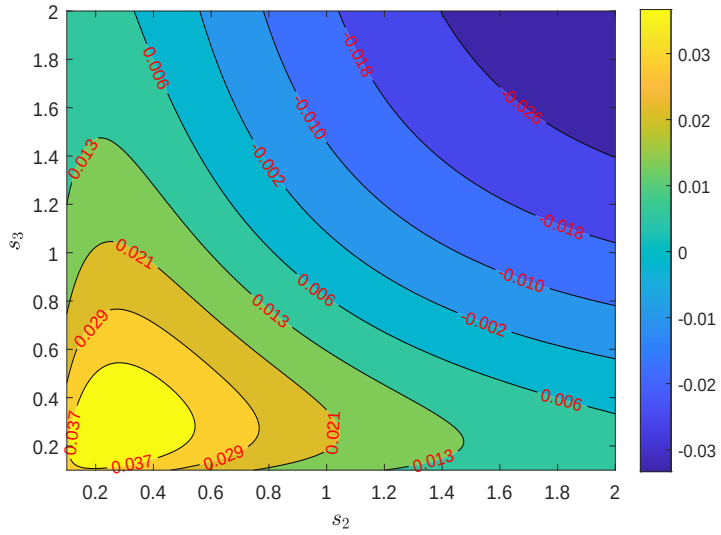


Figure S.1: Contour plot of $P^{\text{SEVI}}(Y = 1) - P^{\text{LEVI}}(Y = 1)$ against (s_2, s_3), the scale values of alternatives 2 and 3 for a fixed $s_1 = 1$

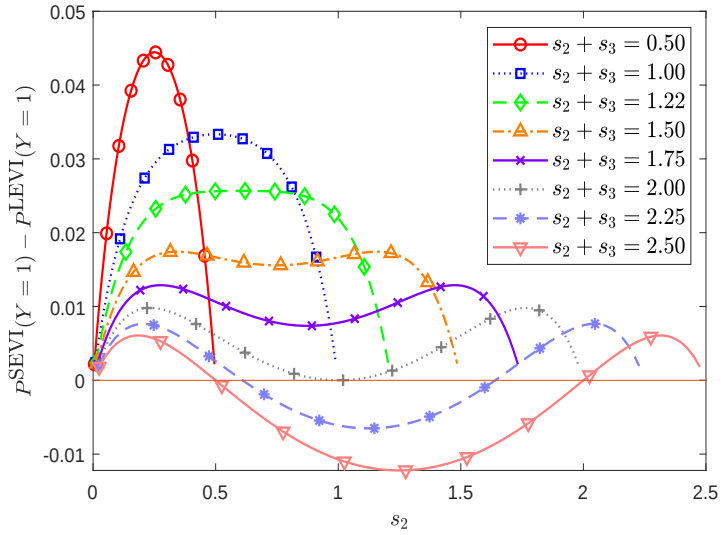


Figure S.2: Plot of $P^{\text{SEVI}}(Y = 1) - P^{\text{LEVI}}(Y = 1)$ against s_2 while holding $s_2 + s_3$ constant for a fixed $s_1 = 1$

Figures S.1 and S.2 are motivated by the following empirical inquiry: Imagine a market with three firms, where firm 1 is the market leader with a scale value $s_1 = 1$ and firms 2 and 3 have a combined scale value $s_2 + s_3$ of 1/2. Under the LEVI assumption, firm 1 dominates the market with a market share of 2/3, while firms 2 and 3 together hold only 1/3. The question is: are consumers more inclined to purchase from the market leader when the other two firms have identical market power (i.e., $s_2 = s_3 = 0.25$), or when one of them dominates the other (e.g., $s_2 = 0.4$ and $s_3 = 0.1$)?

The answer depends on the model used. The standard LEVI logit provides the same prediction for the dominant firm in both scenarios of competitor configurations, because in the logit case only the sum of the competitors' scale values matters: firm 1 holds a share of 2/3 whether its rivals are $(0.25, 0.25)$ or $(0.4, 0.1)$. The SEVI model separates the two. Against evenly matched rivals, firm 1 holds 0.711, about 4.4 percentage points above the LEVI prediction. When one rival dominates the other, it holds 0.697, or 3.1 percentage points above the LEVI prediction. The market leader therefore captures 1.4 percentage points more when facing two equal rivals than when facing one strong and one weak rival, a distinction the logit cannot express at all. This gap between the LEVI and SEVI choice probabilities for the dominant firm narrows as the scale values of the other two firms diverge; see Figure S.2.

S.G.2 The largest rival advantage

Whether an alternative, say alternative 1, is chosen is settled by a single random quantity: the largest advantage that any rival draws over it. Define

$$M := \max_{j=2,\ldots,J} (\varepsilon_j - \varepsilon_1).$$

Alternative 1 is chosen if and only if $v_1 + \varepsilon_1 > v_j + \varepsilon_j$ for every $j \neq 1$, that is, if and only if $\varepsilon_j - \varepsilon_1 < v_1 - v_j$ for every such j. When the $J-1$ rivals share a common systematic utility $v_1 - d$, these $J-1$ inequalities collapse into the single event $\{M < d\}$, so that

$$P(Y = 1|v) = F_M(d), \tag{S.22}$$

where F_M denotes the CDF of M. Evaluated at the systematic lead d, this CDF is therefore the choice probability. Comparing the SEVI and LEVI models under the systematic utilities $(v_1, v_1 - d, \ldots, v_1 - d)$ amounts to comparing F_M across the two distributions of the random component.

Both CDFs are available in closed form. Write $s = e^{-d}$. Under the LEVI, (5) gives

$$F_M^{\text{LEVI}}(d) = \frac{1}{1 + (J-1)\,s} = \frac{1}{1 + e^{-(d - \ln(J-1))}},$$

while (3) with $s^i_{1|k} = 1/s$ for every k gives

$$F_M^{\text{SEVI}}(d) = \int_0^1 \left(1 - u^{1/s}\right)^{J-1} du = \prod_{m=1}^{J-1} \frac{m}{m+s},$$

where the second equality follows from the substitution $t = u^{1/s}$, which turns the integral into $sB(s, J)$. The two forms treat the rivals differently: the LEVI collapses them into a single sum, while the SEVI keeps one factor per rival inside the integral.

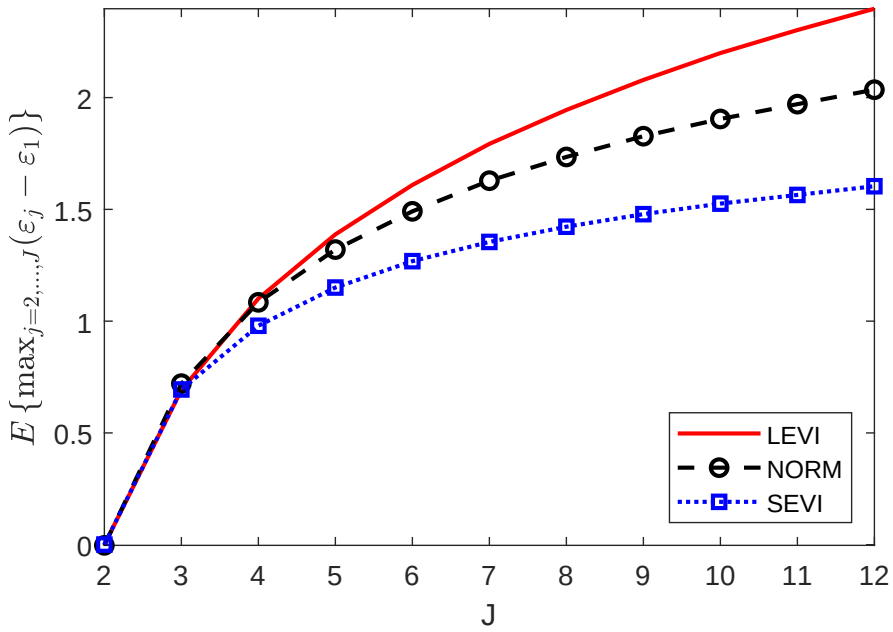


Figure S.3: The mean of the largest advantage $\max_{j=2,\ldots,J}(\varepsilon_j - \varepsilon_1)$ that a rival draws over a fixed reference alternative, under LEVI, $\mathcal{N}(0, \pi^2/6)$, and SEVI

At $d = 0$ so that $s = 1$, both expressions equal $1/J$. Applied with $J - 1$ tied rivals, Theorem 1 shows that for $J \geq 3$ the two CDFs cross there and nowhere else, F_M^{SEVI} lying above F_M^{LEVI} for $d > 0$ and below it for $d < 0$; for $J = 2$ they coincide at every d, both equal to $1/(1 + s)$. That single crossing is "Leader-Takes-More" and "Laggard-Takes-Less" restated in terms of the random component alone: M is less dispersed about zero under the SEVI, so a rival's best draw is less likely either to overturn a lead or to rescue a deficit. The comparison of means that follows is not implied by Theorem 1 and is the substantive content of this subsection.

Figure S.3 plots $E(M)$ against J when $\{\varepsilon_j\}_{j=1}^J$ is iid LEVI, $\mathcal{N}(0, \pi^2/6)$, or SEVI. Under the LEVI, F_M^{LEVI} is the standard logistic CDF centered at $\ln(J - 1)$, so $E(M) = \ln(J - 1)$. The three curves start together at $J = 2$, where M is the difference of two draws and has mean zero under each of the three distributions, and the LEVI and SEVI curves meet again at $J = 3$, where both means can be shown to be $\ln 2$.[15] From $J \geq 4$ onwards the ordering is strict: $E(M)$ is largest under the LEVI, intermediate under the normal distribution, and smallest under the SEVI, and the spread widens with J. Since M is what a systematic lead must overcome, a smaller M means that the random components are less able to overturn the ranking set by the systematic utilities. Holding the systematic utilities fixed, the random components under the LEVI therefore play a larger role in determining the winner than under the SEVI.

[15] This follows from the reflection $\varepsilon^{\text{SEVI}} \stackrel{d}{=} -\varepsilon^{\text{LEVI}}$. Let $\eta_1, \ldots, \eta_J$ be iid LEVI draws. Then $M^{\text{LEVI}} = \max_{j \neq 1} \eta_j - \eta_1$ while $M^{\text{SEVI}} = \eta_1 - \min_{j \neq 1} \eta_j$, so the two means differ by $E(\max_{j \neq 1} \eta_j) + E(\min_{j \neq 1} \eta_j) - 2\gamma$. With $J - 1 = 2$ rivals the elementary identity $\max(a, b) + \min(a, b) = a + b$ forces this difference to vanish, and both means equal $\ln 2$.

S.H Additional Simulation Evidence

This subsection collects the numerical evidence referred to in Sections 2, 3.3, and 5.3. In the first two parts, the data are generated from the two data-generating processes described at the beginning of Section 5: the attributes $X_{ij,\ell}$ are iid $\mathcal{N}(0, \pi^2\omega_j^2/36)$ in the first and iid $\mathcal{N}(0, \pi^2/36)$ in the second, over $\ell = 1, 2, 3$, with $\beta_0 = (1, 2, 1)'$. The third part uses a design of its own, described there.

S.H.1 Choice probabilities

To illustrate the divergence between the SEVI and LEVI models, we simulate the choice probabilities using the linear specification $V_{ij}(\beta_0) = X_{ij}\beta_0$ with iid SEVI or LEVI errors.

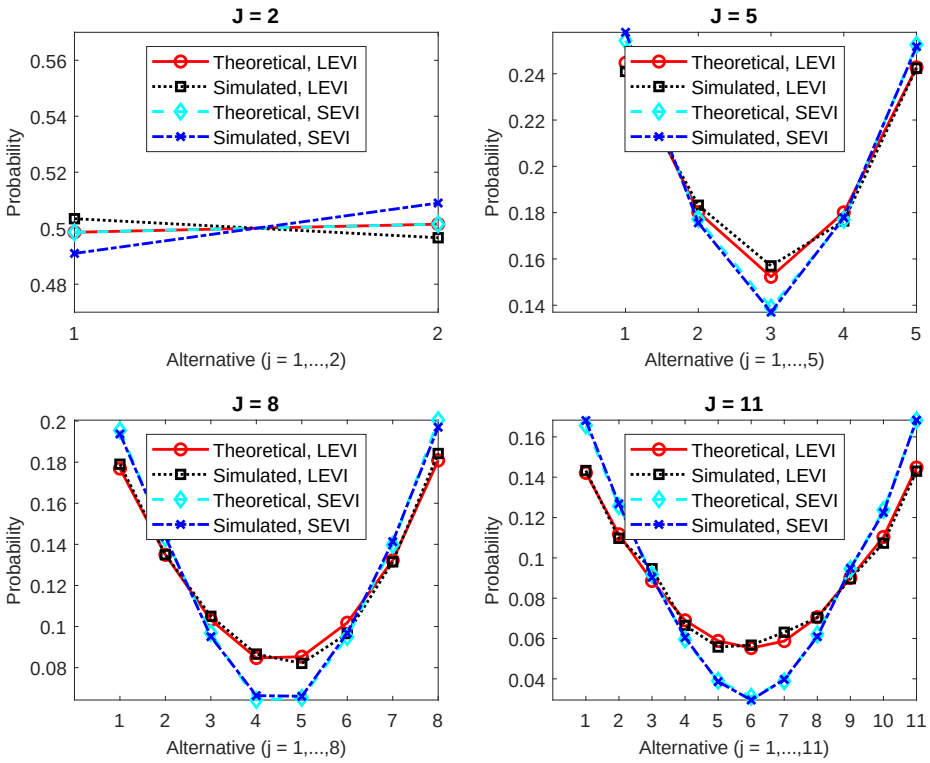


Figure S.4: Simulated choice probabilities and the average of theoretical choice probabilities against alternative labels ($j = 1, \ldots, J$) for a sample of $n = 10,000$ individuals when $X_{ij,\ell}$ is iid $\mathcal{N}(0, \pi^2\omega_j^2/36)$

For each model, we compute the choice probabilities according to the theoretical formulae (i.e., (4) under the SEVI and the standard logit formula (5) under the LEVI), and take the average of these choice probabilities over 10,000 individuals. We also simulate the decisions of 10,000 independent individuals and compute the proportion of individuals choosing each alternative. Figure S.4 reports results for various J values for the first DGP ($X_{ij,\ell} \sim$ iid $\mathcal{N}(0, \pi^2\omega_j^2/36)$). Not surprisingly, since the sample size is large, the average theoretical choice probabilities match the simulated choice proportions well for both the SEVI and LEVI models. When $J = 2$, the theoretical choice probabilities are the same across the two models. With only two alternatives, the simulated choice shares are very close across the two models, differing only slightly due to

simulation errors. When $J > 2$, for alternatives with relatively higher choice probabilities under the LEVI, the SEVI model assigns a higher choice probability than the LEVI model. On the other hand, when $J > 2$, for alternatives with a relatively lower choice probability under the LEVI, the SEVI model assigns a lower choice probability than the LEVI model. Over the range of J considered here, the difference between the choice probabilities across the two models grows with J.

Next, we examine the choice probabilities for the second DGP ($X_{ij,\ell} \sim$ iid $\mathcal{N}(0, \pi^2/36)$) using a different visualization. Figure S.5 provides a scatter plot of the ratio of the choice probabilities in the SEVI and LEVI models (i.e., $P^{\text{SEVI}}/P^{\text{LEVI}}$) against P^{LEVI}. Instead of examining the average choice probabilities as done in Figure S.4, we consider the choice probability of choosing a specific alternative (here, we use alternative 1 without loss of generality) for each realization of the systematic utilities. The panel for $J = 2$ shows the ratio pinned at 1, the coincidence of the two models noted above. Figure S.6 plots the two choice probabilities against each other, together with the 45° line, over a low and a high range of P^{LEVI}. It therefore displays the sign and the absolute size of the difference in the region where the ratio in Figure S.5 is compressed.

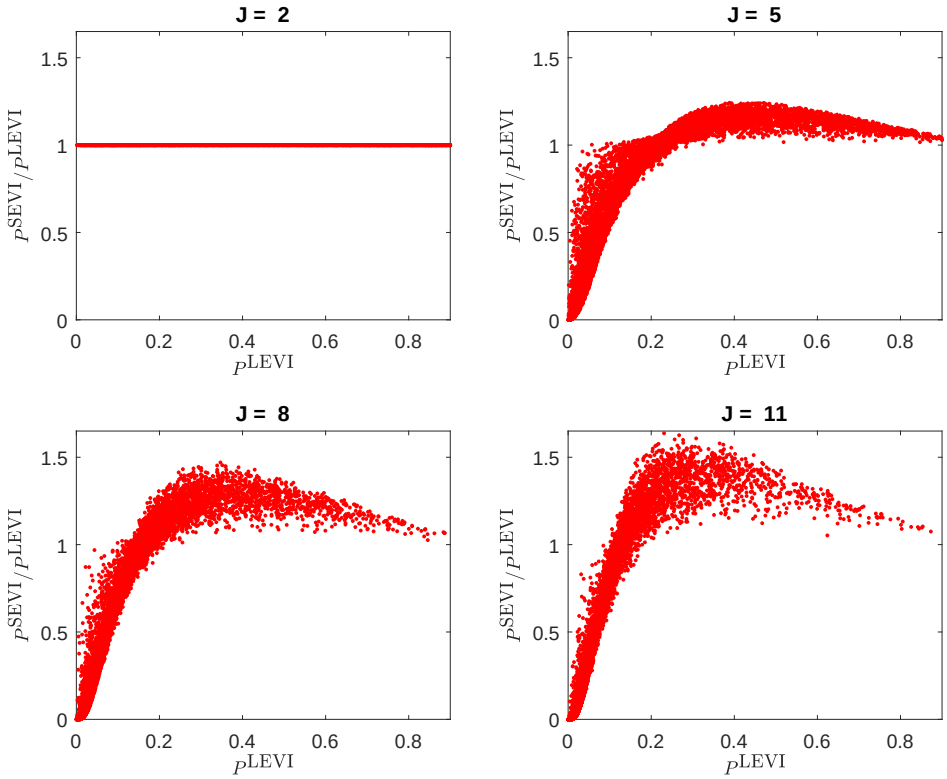


Figure S.5: Scatter plot of the ratio of choice probabilities in the SEVI and LEVI models ($P^{\text{SEVI}}/P^{\text{LEVI}}$) against the choice probability in the LEVI model (P^{LEVI}) with the systematic utilities held the same across the two models when $X_{ij,\ell}$ is iid $\mathcal{N}(0, \pi^2/36)$

In both Figure S.5 and Figure S.6, the systematic utilities are held constant for each data point with either LEVI errors or SEVI errors. We note from these two figures that when $J > 2$ and P^{LEVI} is low, P^{SEVI} can be up to 50% lower: when there are five alternatives ($J = 5$) and $P^{\text{LEVI}} = 0.1$, P^{SEVI} can be as low as 0.05. Thus, from a practical point of view, the choice between the two models can have a substantial impact on predicting the lift for a minor brand

with a market share between 5% and 10%. On the other hand, when $J > 2$ and P^{LEVI} is high, P^{SEVI} tends to be higher. For example, when $J = 5$ and $P^{\text{LEVI}} = 0.4$, P^{SEVI} can be 20% higher; when $J = 11$ and $P^{\text{LEVI}} = 0.3$, P^{SEVI} can be 50% higher. At a given value of P^{LEVI}, the points form a band rather than a curve. When $J = 11$ and $P^{\text{LEVI}} = 0.2$, for instance, the ratio ranges from about 0.9 to about 1.6 across realizations. The size of the difference therefore depends on the whole configuration of the systematic utilities and cannot be read off the LEVI probability alone. The patterns we observe here are similar to those in Figure 1. The difference in the choice probabilities across the two models can be important in various contexts, ranging from antitrust actions to production planning.

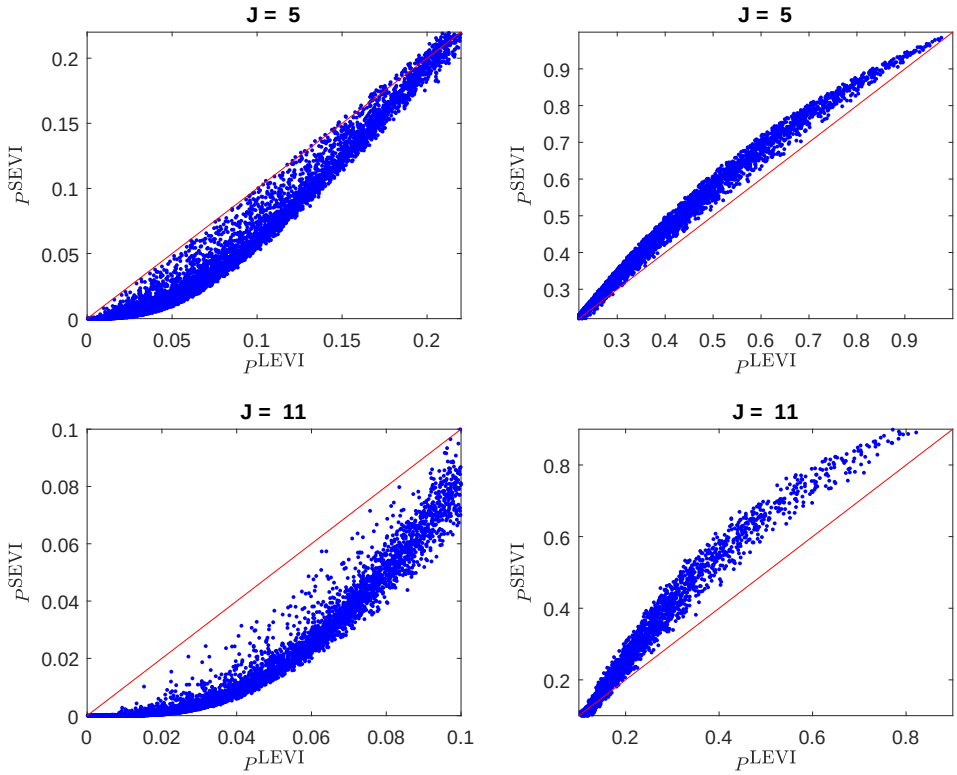


Figure S.6: Scatter plot of the choice probability in the SEVI model versus that in the LEVI model with left and right panels having different ranges for the horizontal axis ($X_{ij,\ell}$ is iid $\mathcal{N}(0, \pi^2/36)$)

The qualitative patterns observed in Figures S.4–S.6 illustrate the "Leader-Takes-More" behavior of SEVI established in Theorem 1(a), which Theorem 2 extends to every threshold at once. The figures also reach beyond both results, at the intermediate ranks where, as explained in Section 2, neither theorem signs the gap between the two models. The tendency recorded by the band above—gain at high LEVI probabilities, loss at low ones—persists through those ranks, and it pushes the SEVI probability vector closer to the one-hot encoding of $\arg\max(v_1, \ldots, v_J)$ than its LEVI counterpart.[16] Section S.G.1 above takes up the case $J = 3$, where the sign and the magnitude of the difference can both be characterized in closed form and neither tendency need hold. In particular, the alternative with the second-highest systematic utility can command a

[16] For ease of discussion, we assume that there is a unique maximum among $v_1, \ldots, v_J$. The 0-1 vector representation of arg max is the so-called one-hot encoding of arg max.

larger market share under the SEVI than under the LEVI even when it falls substantially behind the leader, provided its systematic utility exceeds the average of the other two.

S.H.2 Compensating variations

Outside the laggard cone and the dominance limit of Proposition 3(d), an analytical comparison is difficult, so we compute the ratio of $\mathrm{CV}^{\mathrm{SEVI}}(k, S)$ to $\mathrm{CV}^{\mathrm{LEVI}}(k, S)$ numerically. Figure S.7 plots this ratio against $\mathrm{CV}^{\mathrm{LEVI}}(k, S)$ for $S = \exp(V)$ when V follows the same data-generating process as that in Figure S.4, with one of the covariates in X regarded as the cost. Each point in the figure represents the ratio for a particular instance of k and S. Qualitatively, the figure resembles Figure S.5, where the general patterns are applicable to the DGP in Figure S.4. Specifically, when the CV under the LEVI model is low, the CV under the SEVI model tends to be even lower, and conversely, when the CV under the LEVI model is high, the CV under the SEVI model tends to be higher. The two ends of this pattern are not merely intuitive; they are Proposition 3(b), (c), and (d). A low $\mathrm{CV}^{\mathrm{LEVI}}$ signals that the excluded alternative is a systematic laggard, and there Proposition 3(b) and (c) place the SEVI value strictly below; a high one signals dominance, and there Proposition 3(d) reverses the ranking once $J \geq 4$. What the figure adds is the passage between these two regimes, which none of these results describes: it shows that the ratio may cross one when the alternatives excluded are not the extreme ones.

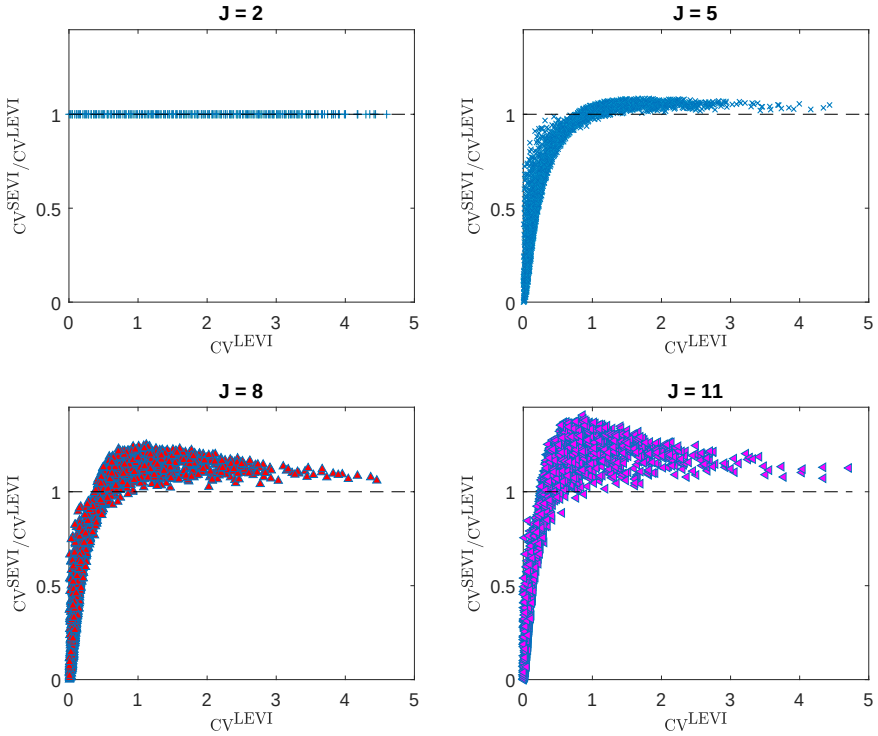


Figure S.7: Scatter plot of the ratio of compensating variations in the SEVI and LEVI models ($\mathrm{CV}^{\mathrm{SEVI}}(k, \exp(V))/\mathrm{CV}^{\mathrm{LEVI}}(k, \exp(V))$) against the compensating variation in the LEVI model ($\mathrm{CV}^{\mathrm{LEVI}}(k, \exp(V))$) with the same DGPs as in Figure S.4 ($X_{ij,\ell}$ is iid $\mathcal{N}(0, \pi^2\omega_j^2/36)$)

S.H.3 Computation time comparison

This part gives the designs, the implementation, and the two figures behind the computation time comparison summarized in Section 5.3.

The data are generated according to $Y_i = \arg\max_{j=1,\ldots,J} \{X_{ij}\beta_0 + \varepsilon_{ij}\}$ where $X_{ij,\ell} \sim iid$ $\mathcal{N}(0, \pi^2\omega_j^2/36)$ over $\ell = 1, \ldots, L$ (cf. equation (26)) and $\varepsilon_{ij} \sim iid$ SEVI over j. We fit the LEVI, SEVI, and NORM models to the so-generated data. All models have the same correct linear specification $V_{ij}(\beta) = X_{ij}\beta$ for the systematic utility but differ with respect to the distributional specification of the random components.

We use a sample size of 1,000, explore different numbers of alternatives (J) ranging from 2 to 16, and consider two different numbers of attributes ($L = 3$ and $L = 10$). To mimic real data analysis, we simulate only one sample with model parameter β_0 drawn randomly from a multivariate standard normal distribution.

We document the time needed to obtain the ML/QML estimate of the model parameters for each of the three models. The time represents the seconds spent by our MATLAB program to find each point estimate. All computations are performed on a desktop PC equipped with an Intel i7-13700F processor and 32GB of RAM. While no parallel method is employed,[17] we strive to vectorize all operations as much as possible. For the SEVI model, the choice probabilities are computed using the closed formula in Lemma 1, with the choice subsets enumerated by Gosper's hack (Knuth, 2011, §7.1.3, Exercise 20), which produces each successive subset from the previous one at constant cost.[18] In the case of the NORM model, the choice probabilities are obtained using the Geweke–Hajivassiliou–Keane (GHK) simulator with 500 simulations (Geweke and Keane, 2001; Hajivassiliou and McFadden, 1998; Keane and Wolpin, 1994).

Figure S.8 plots the computation time against the number of alternatives. On the one hand, when $J \leq 12$, the computation time for the SEVI model is comparable to that for the LEVI model. When $J > 12$, the computation time for the SEVI model starts to increase with J and becomes significantly larger than that for the LEVI model, especially when J reaches 16. On the other hand, when $J \leq 12$, the NORM model exhibits a clear computational disadvantage compared to the LEVI and SEVI models. While the computation time for each of the latter two models is less than 27 seconds when the number of attributes is 10 and the number of alternatives is 12, the corresponding computation time for the NORM model is more than 1,000 seconds. In our experiment, the SEVI model starts to become as computationally costly as the NORM model only when $J = 16$.

While the DGP just described contains no alternative-specific constants, the second group of DGPs does. We examine two scenarios: one with only alternative-specific constants where $V_{ij} = Z_i\beta_{zj}$ for $Z_i = 1$ (hereafter, ASC-only) and another with alternative-specific constants (Z_i) and three attributes $X_{ij,\ell} \sim iid$ $\mathcal{N}(0, \pi^2\omega_j^2/36), \ell = 1, 2, 3$. In the latter case, $V_{ij} = Z_i\beta_{zj} + X_{ij,1}\beta_{x1} + X_{ij,2}\beta_{x2} + X_{ij,3}\beta_{x3}$ where $Z_i = 1$. In both cases, $\varepsilon_{ij} \sim iid$ SEVI over $j = 1, \ldots, J$.

[17] This represents one potential direction to explore for improving the computational speed of the SEVI model for larger values of J. A second is to bypass the closed form. When J is large, evaluating the integral representation in Lemma 1 by quadrature can be cheaper than summing its terms, though no single rule performs well across all configurations. The computation times reported here are therefore not the best attainable.

[18] The number of terms in (4) is $2^{J-1} - 1$, so evaluating it directly is a combinatorial problem that grows quickly with J. The hack encodes a subset $\mathcal{S} \subseteq \mathcal{J}_{-j}$ as a $(J-1)$-bit word whose ones mark the members of $\mathcal{S}$, and advances from any such word to the next larger word with the same number of ones in a fixed number of bit operations. Iterating from the smallest word with ℓ ones therefore visits each of the $\binom{J-1}{\ell}$ subsets in the inner sum exactly once, without holding the enumeration in memory. Because (4) is exact, no simulation is required; the enumeration delivers computational efficiency at moderately large choice sets.

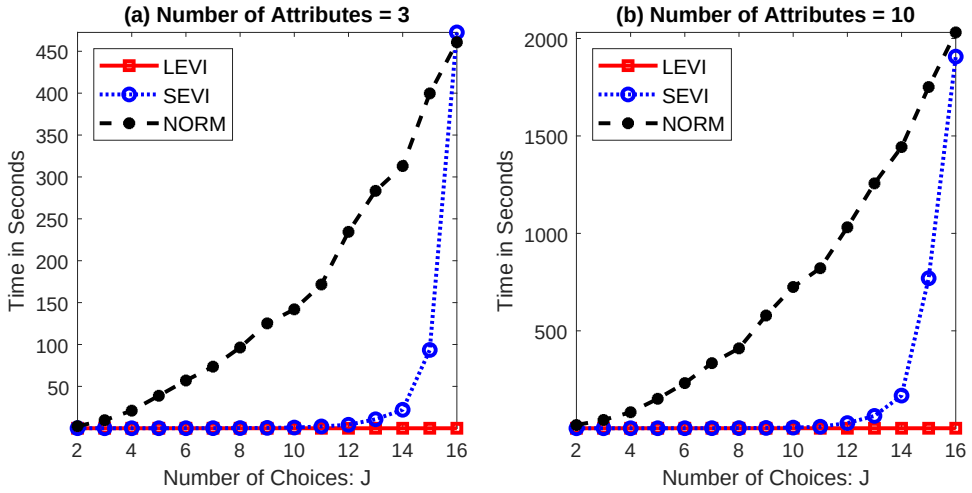


Figure S.8: Comparison of computation times for the LEVI, SEVI, and NORM (MNP with $\varepsilon_{ij} \sim iid\mathcal{N}(0, \pi^2/6)$) models when $V_{ij} = X_{ij}\beta$

For each dataset, we fit the LEVI, SEVI, and NORM models with correctly specified systematic utility. Figure S.9 reports the computation times. The general patterns observed in Figure S.8 carry over to this figure. When $J \leq 12$, the SEVI and LEVI models require less than 65 seconds for estimation, with much shorter times for a smaller J. In contrast, the time needed for the NORM model grows rapidly as J increases from 2 to 12. For example, when $J = 10$ and 12, the required times for the NORM model for the ASC-only case are 2,274 and 5,159 seconds, respectively. In other words, estimating the NORM model takes between half an hour and two hours when J is about 10 or 12, whereas the corresponding time for a SEVI model is only about 1 minute. As before, the SEVI model becomes computationally less competitive for $J > 12$ and has a computational cost starting to approach that of the NORM model when J reaches 16.

S.I Supplementary Figures

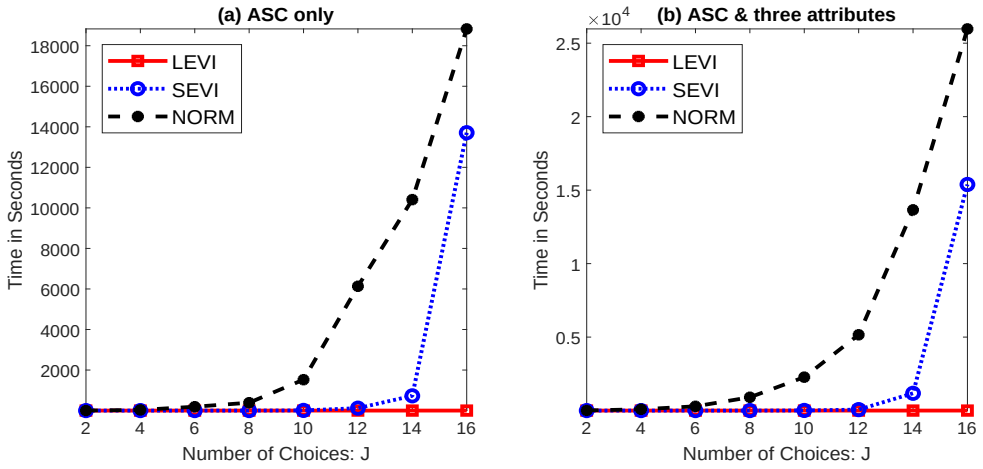


Figure S.9: Comparison of computation times for the LEVI, SEVI, and NORM (MNP with $\varepsilon_{ij} \sim iid\mathcal{N}(0, \pi^2/6)$) models when $V_{ij} = Z_i\beta_{zj}$ or $V_{ij} = Z_i\beta_{zj} + X_{ij}\beta_x$

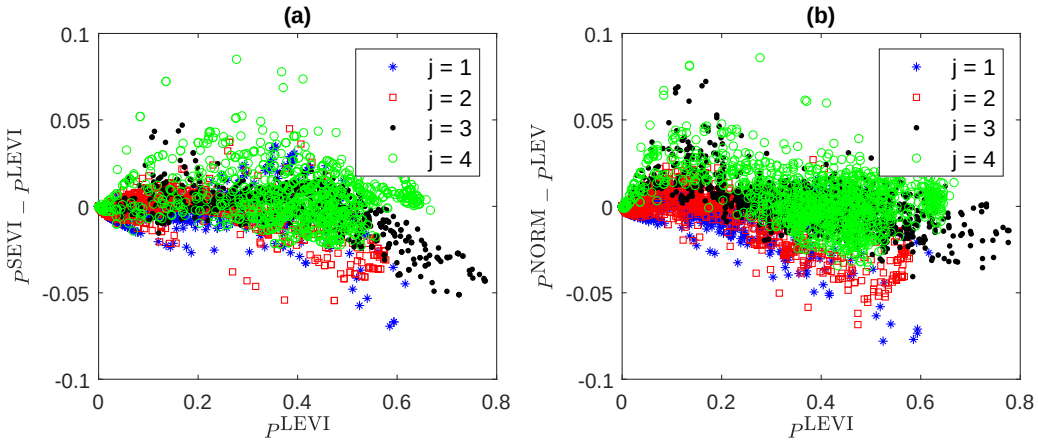


Figure S.10: Plot of the difference of the predicted choice probabilities against the LEVI predicted probability in the fishing mode choice application

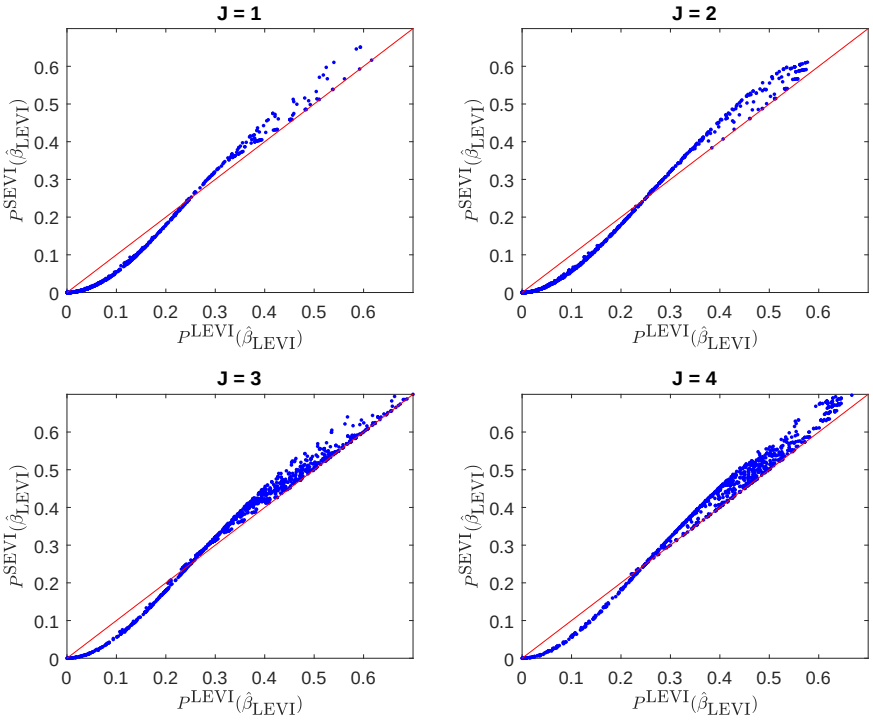


Figure S.11: Scatter plot of $P^{\text{SEVI}}(\hat{\beta}_{\text{LEVI}})$ against $P^{\text{LEVI}}(\hat{\beta}_{\text{LEVI}})$ based on the fishing mode choice data

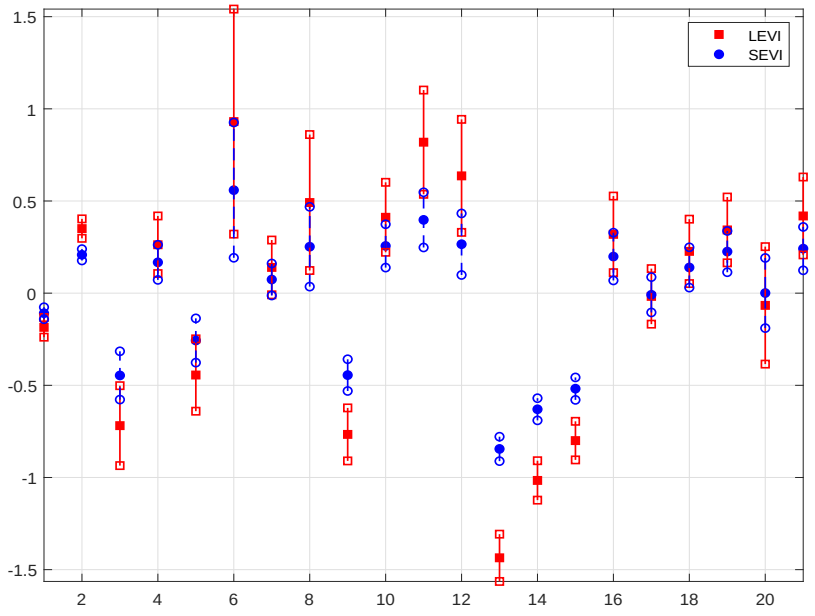


Figure S.12: Point estimates and 95% confidence intervals for the 21 parameters in the LEVI and SEVI models for the vehicle choice application

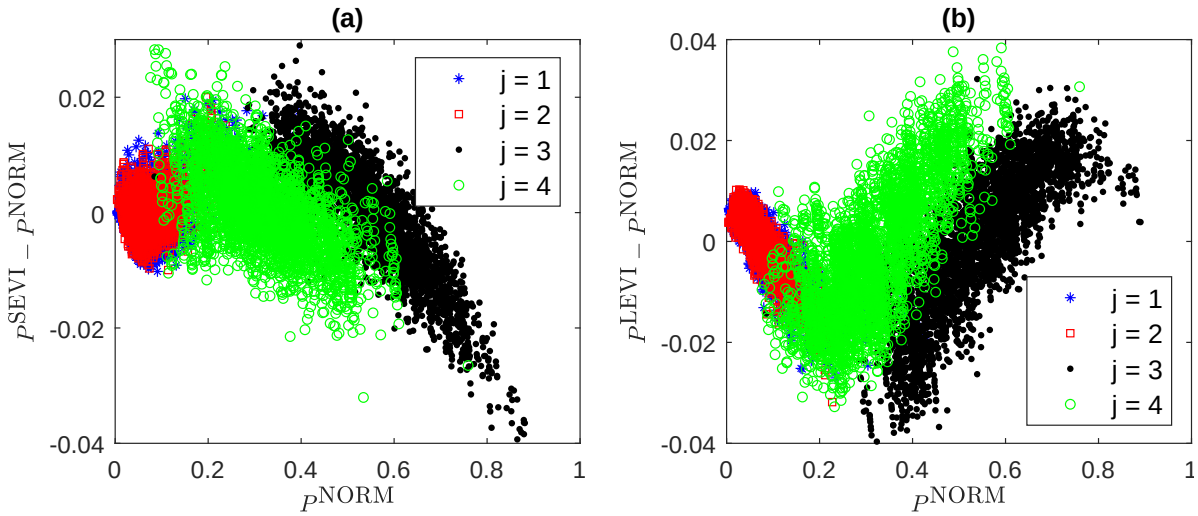


Figure S.13: Plot of the difference of the predicted choice probabilities against the NORM predicted probability in the cracker choice application

References

Anderson, S. P. and de Palma, A. (1999). Reverse discrete choice models. *Regional Science and Urban Economics*, 29(6):745–764.

Barberá, S. and Pattanaik, P. K. (1986). Falmagne and the rationalizability of stochastic choices in terms of random orderings. *Econometrica*, 54(3):707–715.

Brownstone, D., Bunch, D. S., Golob, T. F., and Ren, W. (1996). A transactions choice model for forecasting demand for alternative-fuel vehicles. *Research in Transportation Economics*, 4:87–129.

Brownstone, D. and Train, K. (1998). Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics*, 89(1):109–129.

Cameron, A. C. (2022). *Analysis of Economics Data: An Introduction to Econometrics*. Amazon.com.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

Cameron, A. C. and Trivedi, P. K. (2022). *Microeconometrics using Stata*. Stata Press, College Station, Texas.

Collins, A. T. and Rose, J. M. (2013). Estimation of stochastic scale with best-worst data. Technical Report ITLS-WP-13-13, Institute of Transport and Logistics Studies, University of Sydney.

Croissant, Y. (2020). Estimation of random utility models in R: the mlogit package. *Journal of Statistical Software*, 95(11):1–41.

Daly, A. and Zachary, S. (1978). Improved multiple choice models. In Hensher, D. A. and Dalvi, M. Q., editors, *Identifying and Measuring the Determinants of Mode Choice*. Teakfield, London.

de Haan, L. and Ferreira, A. (2007). *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer New York.

de Palma, A. and Kilani, K. (2023). Best, worst, and best&worst choice probabilities for logit and reverse logit models. *Journal of Choice Modelling*, 49:100449.

Delle Site, P., Kilani, K., Gatta, V., Marcucci, E., and de Palma, A. (2019). Estimation of consistent logit and probit models using best, worst and best–worst choices. *Transportation Research Part B: Methodological*, 128:87–106.

Fowlie, M. (2010). Emissions trading, electricity restructuring, and investment in pollution abatement. *American Economic Review*, 100(3):837–69.

Geweke, J. and Keane, M. P. (2001). Computationally intensive methods for integration in econometrics. In Heckman, J. J. and Leamer, E. E., editors, *Handbook of Econometrics*, volume 5, pages 3463–3568. Elsevier, Amsterdam.

Gumbel, E. J. (1958). *Statistics of Extremes*. Columbia University Press, New York.

Hajivassiliou, V. A. and McFadden, D. (1998). The method of simulated scores for the estimation of LDV models. *Econometrica*, 66(4):863–896.

Hausman, J. and McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica*, 52(5):1219–1240.

Hensher, D. A., Rose, J. M., and Greene, W. H. (2015). *Applied Choice Analysis: A Primer*. Cambridge University Press, 2nd edition.

Herriges, J. A. and Kling, C. L. (1999). Nonlinear income effects in random utility models. *Review of Economics and Statistics*, 81(1):62–72.

Hu, W. (2005). Logit models: smallest versus largest extreme value error distributions. *Applied Economics Letters*, 12(12):741–744.

Jain, D. C., Vilcassim, N. J., and Chintagunta, P. K. (1994). A random-coefficients logit brand-choice model applied to panel data. *Journal of Business & Economic Statistics*, 12(3):317–328.

Keane, M. P. and Wolpin, K. I. (1994). The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte Carlo evidence. *Review of Economics and Statistics*, 76(4):648–672.

Knuth, D. (2011). *The Art of Computer Programming, Volume 4A: Combinatorial Algorithms, Part 1*. Pearson Education, Upper Saddle River, NJ.

Manski, C. F. (1977). The structure of random utility models. *Theory and Decision*, 8(3):229–254.

Marshall, A. W., Olkin, I., and Arnold, B. C. (2011). *Inequalities: Theory of Majorization and Its Applications*. Springer Science & Business Media, New York, second edition.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behaviour. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York, NY, USA.

McFadden, D. (1981). Econometric models of probabilistic choice. In Manski, C. F. and McFadden, D., editors, *Structural Analysis of Discrete Data with Econometric Applications*, pages 198–272. MIT Press.

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

Misra, S. (2005). Generalized reverse discrete choice models. *Quantitative Marketing and Economics*, 3(2):175–200.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In Engle, R. F. and McFadden, D. L., editors, *Handbook of Econometrics*, volume 4, chapter 36, pages 2111–2245. Elsevier, Amsterdam.

Paap, R. and Franses, P. H. (2000). A dynamic multinomial probit model for brand choice with different long-run and short-run effects of marketing-mix variables. *Journal of Applied Econometrics*, 15(6):717–744.

Pólya, G. and Szegő, G. (1976). *Problems and Theorems in Analysis II: Theory of Functions, Zeros, Polynomials, Determinants, Number Theory, Geometry*. Springer-Verlag, Berlin. Part Five, Chapter 1, Problem 75.

Ruud, P. A. (1986). Consistent estimation of limited dependent variable models despite misspecification of distribution. *Journal of Econometrics*, 32(1):157–187.

Strzalecki, T. (2023). *Stochastic Choice Theory*. preprint, Department of Economics, Harvard University. `https://scholar.harvard.edu/sites/scholar.harvard.edu/files/tomasz/files/scnotes94k.pdf`.

Sørensen, J. R.-V. and Fosgerau, M. (2022). How McFadden met Rockafellar and learned to do more with less. *Journal of Mathematical Economics*, 100:102629.

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

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57(2):307–333.

Williams, H. (1977). On the formation of travel demand models and economic evaluation measures of user benefit. *Environment & Planning A: Economy and Space*, 9(3):285–344.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge and London.

Yellott, J. I. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, 15(2):109–144.