ISSN 1745-8587

Birkbeck UNIVERSITY OF LONDON

Department of Economics, Mathematics and Statistics

Birkbeck Working Papers in Economics & Finance

BWPEF 1512

Closed Form Solutions for the Generalized Extreme Value Distribution

Walter Beckert
Birkbeck, University of London

Yuya Takahashi
Johns Hopkins University

November 2015

Closed Form Solutions for the Generalized Extreme Value Distribution*

Walter Beckert[1] and Yuya Takahashi[2]

[1]Birkbeck College, University of London
[2]Johns Hopkins University

November 12, 2015

Abstract

The analytical tractability and its maximum stability property make the generalized extreme value (GEV) distribution an attractive choice in the theoretical and econometric modelling of unobservables in incomplete information games. This paper presents new results on conditional moments of order statistics of GEV distributed random variables. And it provides a recursive algorithm to derive the GEV density in high dimensional problems, thereby enabling simulating the Nested Multinomial Logit (NMNL) model on the basis of the Markov chain Monte Carlo protocol of McFadden (1999).

Key words: generalized extreme value distribution, order statistics
JEL classification: C15, C25, C46, C57

*Thanks to Arina Nikandrova for helpful comments.

1 Introduction

Order statistics play important roles in theoretical and applied microeconomic analysis, typically in a setting of incomplete information. Conditional moments can serve to impute unobservables, given relevant information. Closed form expressions for them are attractive in theoretical work because they facilitate, or indeed enable, further analytical analysis, such as comparative statics. And they are useful in empirical work because they avoid costly simulation and render moment based estimation computationally more efficient. In applied demand analysis, the analytical tractability of the Generalized Extreme Value (GEV) distribution has made it a popular modelling choice, giving rise to the family of models involving logit choice probabilities[1]. This paper presents novel analytical results on conditional distributions and moments of order statistics of GEV distributed random variables.

Decisions under uncertainty, made by "Bayesian-Nash" players, are intrinsic to static and dynamic games of incomplete information. Typical examples of applied interest falling into this class are auctions[2] and models of multilateral bargaining[3]. Conditional moments of order statistics are thereby integral to the games' Bayesian-Nash equilibria. In auctions, for example, bidders do not know the valuations of rival bidders, and in a setting of independent private valuations (IPV), the optimal Bayesian-Nash bidding strategy, irrespective of auction mechanism, induces bidding strategies whose payments equal the expectation of the second highest valuation, conditional on one's own bid being the highest[4]. An analogous result holds for auctions with some types of asymmetric bidders[5]. In dynamic bargaining with incomplete information, the analysis in extensive form games requires a notion of beliefs about ex ante unknown gains from trade at each information set of the game, i.e. a consistent conditional probability distribution, and its implied conditional

[1]McFadden (1974, 1981), and the huge literature following these papers.

[2]See Athey and Haile (2007) for an overview.

[3]Ausubel, Cramton and Deneckere (2002); see also, e.g., Fudenberg and Tirole (1983), Chatterjee and Samuelson (1983), Myerson (1984); see also the related literature on mechanism design, e.g Myerson and Satterthwaite (1983).

[4]Vickrey (1961), Myerson (1981), Samuelson and Riley (1981).

[5]Maskin and Riley (2000), Froeb, Tschanz and Crooke (1998); also Brannman and Froeb (2000).

expectation, of reaching that information set and of the gains flowing from it[6].

The GEV distribution is a convenient modelling choice for unobservable or uncertain payoffs in static games or mechanisms where agents' optimal Bayesian-Nash strategies involve considerations about the maximum payoff to rival agents; and in dynamic mechanisms where an agent at any stage in the sequence considers the maximum payoff of subsequent episodes. To the extent that those payoffs across agents and episodes are independent, their maximum is also distributed EV type 1. The type 1 EV (Gumbel) distribution thus satisfies the maximum stability postulate[7]. As it turns out, the GEV distribution enjoys great analytical tractability. The convenience of analytical tractability, paired with maximum stability, make the GEV distribution an attractive choice to model incomplete information structures.

Analytical tractability entails the additional benefit in that it permits closed form expressions for comparative static results, e.g. characterizing the effect of an increase in the number of bidders on the expected revenue in an IPV auction. Such results, next to the well-known analytical expressions for own and cross elasticities in logit demand models, are of great value, e.g. in applied competition analysis. In the microeconometric analysis of incomplete information games, the data typically only capture the value of the winning agents' optimal strategies, e.g the winning bid or the price reached in the final of a sequence of bargaining episodes. The values of rival agents' strategies along and off the equilibrium path, such as losing bids and inferior bargaining matches, typically are not observed. To the extent that agents' optimal strategies are constrained by, and hence depend on, such values, structural econometric analysis needs to properly account for them. This can be done relatively efficiently if they can be replaced - or imputed - by expectations, conditional on observables; see, for example, Beckert, Smith and Takahashi (2015). An analytical E (or expectation) step does not only avoid additional computations necessary in simulation and numerical approximations, but it also improves the precision of resulting estimators relative to their simulation-assisted counterparts[8].

---

[6] Crampton (1985).

[7] David(1981); the Fréchet or EV type 2 distribution is the only other distribution satisfying this postulate.

[8] McFadden (1989) and Pakes and Pollard (1989) provide results on the additional imprecision of estimators due to simulation. Typically, there is a trade-off between estimator precision and

This paper provides novel results on conditional distributions and moments of order statistics of GEV distributed random variables. These are complemented by a result that establishes an algorithm to derive the GEV density in high dimensional nested multinomial logit problems. This algorithm is critical to operationalize the Markov chain Monte Carlo (MCMC) sampling scheme proposed by McFadden (1999) to obtain random draws from GEV distributions in order to simulate high dimensional nested logit models. Such models are common in the differentiate products literature where the number of choice alternatives can rapidly exceed 100 (Berry et al. (1995), Beckert at al. (2015)). Simulating these models is of interest, for instance, for the purpose of evaluating counterfactuals such as hypothetical mergers and introductions of new products, or, as in McFadden (1999), to approximate mean willingness-to-pay measures in GEV models that are nonlinear in income.

2 Framework

This section sets out the modeling framework within which the subsequent results can be considered. For future reference, it provides two auxiliary results: It shows (i) the well-known maximum stability property that the maximum order statistic of conditionally extreme value distributed surpluses has also an extreme value distribution, and (ii) that a restricted, two-stage nested optimization algorithm in an EV model is consistent with an EV model on the second stage.

Consider the indirect utility, or surplus, accruing to a decision maker arising from choice alternative k, denoted by S_k, $k \in \mathcal{K} = \{1, \cdots, K\}$. Let $S_k = \delta_k + \sigma\epsilon_k$, where $\delta_k \in \mathbb{R}$ is a location parameter, $\sigma > 0$ is a scale parameter, and ϵ_k is an i.i.d. extreme value type 1 residual, i.e. $\epsilon_k \overset{i.i.d.}{\sim} EV(0, 1)$.

Result (i): *Distribution of the maximum order statistic.*

For $y \in \mathbb{R}$, $S_{K:K} := \max\{S_k, k = 1, \cdots, K\}$ is distributed extreme value, with

computational effort: simulation inflates the variance-covariance matrix of estimators inversely proportional to the number of simulation sample draws; the computational cost, however, increases in this number.

location parameter $I = \sigma \ln \left(\sum_{k=1}^{K} \exp\left(\frac{\delta_k}{\sigma}\right) \right)$.

Proof:

$$
\begin{aligned}
F_{S_{K:K}}(y) &= \Pr\left(S_{K:K} < y\right) \\
&= \Pr\left(\delta_k + \sigma\epsilon_k < y, \ \forall \ k = 1, \cdots, K\right) \\
&= \prod_{k=1}^{K} \Pr\left(\epsilon_k < \frac{y - \delta_k}{\sigma}\right) \\
&= \prod_{k=1}^{K} \exp\left(-\exp\left(-\frac{y - \delta_k}{\sigma}\right)\right) \\
&= \exp\left(-\exp\left(-\frac{y}{\sigma} + \ln\left(\sum_{k=1}^{K} \exp\left(\frac{\delta_k}{\sigma}\right)\right)\right)\right) \\
&= \exp\left(-\exp(-\frac{y - I}{\sigma})\right),
\end{aligned}
$$

and hence,

$$
f_{S_{K:K}}(y) = \frac{d}{dy} F_{S_{K:K}}(y) = \exp\left(-\frac{y - I}{\sigma} - \exp\left(-\frac{y - I}{\sigma}\right)\right).
$$

□

Result (ii): *Sequential Optimization.*

Suppose the K choice alternatives can be partitioned such that $k_j \in \mathcal{K}_j$, $j \in \mathcal{J}$, $\mathcal{K}_j \cap \mathcal{K}_l = \emptyset$ if $j \neq l$, and $\bigcup_{j \in \mathcal{J}} \mathcal{K}_j = \mathcal{K}$. Let $K_j = \text{card}(\mathcal{K}_j)$, and $S_{k_j} = \alpha_{k_j} + \sigma\epsilon_{k_j}$, where $\alpha_{k_j} \in \mathbb{R}$ is a location parameter and ϵ_{k_j} is i.i.d. extreme value type 1, $k_j \in \mathcal{K}_j$, $j \in \mathcal{J}$. Assume that optimization proceeds in two steps: (1) $S_{K_j:K_j} = \max\{S_{k_j}, k_j \in \mathcal{K}_j\}$, and (2) $\max\{S_{K_j:K_j}, j \in \mathcal{J}\}$. Then, step (2) is equivalent to an multinomial logit model, with $J = \text{card}(\mathcal{J})$ choice alternatives, whose surpluses are i.i.d. extreme value, with location parameters $\delta_j = \sigma \ln \left(\sum_{k \in \mathcal{K}_j} \exp\left(\frac{\delta_k}{\sigma}\right) \right)$.

Proof: This follows by backwards induction from Result (i) for the step (1) optimization, and the standard logit model for step (2). □

This type of sequential structure is frequently observed in models of multi-product firms (Berry et al. (1995)). Suppose that every oligopolistic firm produces several distinct products. Consumers choose the best alternative, while the price of the best alternative could depend on surplus generated from the best rival product.[9] In this case, the researcher may want to know the expected value of the best rival product given a chosen alternative. The result of sequential optimization is useful to compute such an expected value, as illustrated in the example in Section 4.

3 Some Results on Conditional Distributions and Expectations

Consider the above framework with J choice alternatives, each with surplus $S_j = \delta_j + \sigma\epsilon_j$, $j \in \mathcal{J} = \{1, \cdots, J\}$. Also, for ease of exposition, suppose that $S_1 > S_2 > S_3 > S_k$, $k \geq 4$. Let $\mathcal{S} = \{S_j, j \in \mathcal{J}\}$, and $\mathcal{S}_{-j} = \mathcal{S} \setminus \{j\}$. Furthermore, let $p_j = \frac{\exp\left(\frac{\delta_j}{\sigma}\right)}{\sum_{j\in\mathcal{J}} \exp\left(\frac{\delta_j}{\sigma}\right)}$, $I = \sigma \ln\left(\sum_{j\in\mathcal{J}} \exp\left(\frac{\delta_j}{\sigma}\right)\right)$, and $I_{-j} = \sigma \ln\left(\sum_{\substack{k\in\mathcal{J}\\ k\neq j}} \exp\left(\frac{\delta_k}{\sigma}\right)\right)$. The setup in Beckert et al. (2015), for example, has $J = 4$.

Result 1: *Conditional distribution of the second highest surplus, given the highest surplus.*

Consider $S_{J:J} = \max \mathcal{S}$ and $S_{J-1:J} = \max \mathcal{S} \setminus \{S_{J:J}\} = \max \mathcal{S}_{-1}$. For $y \in \mathbb{R}$,

$$F_{S_{J-1:J}|S_{J:J}}(y|S_{J:J} = S_1) = \exp\left(-\exp\left(-\left(\frac{y-\delta_1}{\sigma} + \ln p_1\right)\right)\right) + \frac{1}{p_1}\left[\exp\left(-\exp\left(-\frac{y-I_{-1}}{\sigma}\right)\right) - \exp\left(-\exp\left(-\frac{y-I}{\sigma}\right)\right)\right].$$

Proof:

$$\Pr\left(\max \mathcal{S}_{-1} < y | S_1 = \max \mathcal{S}\right) = \Pr\left(S_j < y \;\;\&\;\; S_j < S_1, j \in \mathcal{J}_{-1}\right)/p_1.$$

[9] This is due to competition among firms. For the detail, see Beckert et al. (2015).

Consider the numerator:

$$\begin{aligned}
P(y,\delta_1) &= \Pr\left(S_j < y \ \& \ S_j < S_1, j \in \mathcal{J}_{-1}\right) \\
&= \mathbb{E}_{\epsilon_1}\left[\prod_{j\in\mathcal{J}_{-1}} \Pr\left(\epsilon_j < \frac{y-\delta_j}{\sigma}, \epsilon_j < \frac{\delta_1-\delta_j}{\sigma} + \epsilon_1 \middle| \, \epsilon_1\right)\right] \\
&= \mathbb{E}_{\epsilon_1}\left[\prod_{j\in\mathcal{J}_{-1}} \Pr\left(\epsilon_j < \frac{\delta_1-\delta_j}{\sigma} + \min\left\{\frac{y-\delta_1}{\sigma}, \epsilon_1\right\} |\epsilon_1\right)\right] \\
&= \mathbb{E}_{\epsilon_1}\left[\prod_{j\in\mathcal{J}_{-1}} \exp\left(-\exp\left(-\left(\frac{y-\delta_j}{\sigma}\right)\right)\right) 1_{\left\{\frac{y-\delta_j}{\sigma}\leq\epsilon_1\right\}}\right. \\
&\quad \left. + \prod_{j\in\mathcal{J}_{-1}} \exp\left(-\exp\left(-\left(\epsilon_1 + \frac{\delta_1-\delta_j}{\sigma}\right)\right)\right) 1_{\left\{\epsilon_1<\frac{y-\delta_1}{\sigma}\right\}}\right]
\end{aligned}$$

$$\begin{aligned}
&= \exp\left(-\sum_{j\in\mathcal{J}_{-1}} \exp\left(-\left(\frac{y-\delta_j}{\sigma}\right)\right)\right)\left(1-\exp\left(-\exp\left(-\frac{y-\delta_1}{\sigma}\right)\right)\right) \\
&\quad + \int_{-\infty}^{\frac{y-\delta_1}{\sigma}} \exp\left(-\sum_{j\in\mathcal{J}_{-1}} \exp\left(-\left(\epsilon_1 + \frac{\delta_1-\delta_j}{\sigma}\right)\right)\right) \exp(-\epsilon_1 - \exp(-\epsilon_1)) d\epsilon_1 \\
&= \exp\left(-\exp\left(-\frac{y}{\sigma}\right)\left(\sum_{j\in\mathcal{J}_{-1}} \exp\left(\frac{\delta_j}{\sigma}\right)\right)\right)\left(1-\exp\left(-\exp\left(-\frac{y-\delta_1}{\sigma}\right)\right)\right) \\
&\quad + \int_{-\infty}^{\frac{y-\delta_1}{\sigma}} \exp\left(-\epsilon_1 - \exp(-\epsilon_1)\left(1 + \sum_{j\in\mathcal{J}_{-1}} \exp\left(\frac{\delta_j-\delta_1}{\sigma}\right)\right)\right) d\epsilon_1 \\
&= \exp\left(-\exp\left(-\frac{y-I_{-1}}{\sigma}\right)\right) - \exp\left(-\exp\left(-\frac{y-I}{\sigma}\right)\right) \\
&\quad + \int_{-\infty}^{\frac{y-\delta_1}{\sigma}} \exp\left(-\epsilon_1 - \exp(-\epsilon_1 - \ln p_1)\right) d\epsilon_1 \\
&= \exp\left(-\exp\left(-\frac{y-I_{-1}}{\sigma}\right)\right) - \exp\left(-\exp\left(-\frac{y-I}{\sigma}\right)\right) \\
&\quad + p_1 \exp\left(-\exp\left(-\left(\frac{y-\delta_1}{\sigma} - \ln p_1\right)\right)\right).
\end{aligned}$$

As a consistency check:

$$\begin{aligned} \lim_{y \to +\infty} P(y, \delta_1) &= p_1 \\ \lim_{\delta_1 \to +\infty} P(y, \delta_1) &= \exp\left(-\exp\left(-\frac{y - I_{-1}}{\sigma}\right)\right). \end{aligned}$$

□

Result 2: *Conditional mean of the second highest surplus, given the highest surplus.*

$$\mathbb{E}\left[S_{J-1:J} | S_{J:J} = S_1\right] = \sigma\gamma + I + \sigma \ln(1 - p_1)/p_1,$$

where γ is Euler's constant.

Proof: Result 1 implies for the first term that its mean is $\delta_1 - \sigma \ln p_1 + \sigma\gamma$, while it implies for the second term that its mean is $\frac{1}{p_1}\left[I_{-1} - I\right]$. Therefore,

$$\begin{aligned} \mathbb{E}\left[S_{J-1:J} | S_{J:J} = S_1\right] &= \sigma\gamma + \delta_1 - \sigma \ln p_1 + \frac{1}{p_1}\left[I_{-1} - I\right] \\ &= \sigma\gamma + I + \frac{\sigma}{p_1} \ln(1 - p_1), \end{aligned}$$

where $\gamma = \sigma e$. □

Remark: This is the same result as Brannman and Froeb (2000), equation (6), using a result in Froeb et al. (1998).

Let $\mathcal{J}_{-1} = \mathcal{J} \setminus \{1\}$, and define $p_{2|\mathcal{J}_{-1}} = \frac{\exp\left(\frac{\delta_2}{\sigma}\right)}{\sum_{j \in \mathcal{J}_{-1}} \exp\left(\frac{\delta_j}{\sigma}\right)}$.

Result 3: *Probability of first and second highest surplus.*

$$\Pr\left(S_{J:J} = S_1 \;\&\; S_{J-1:J} = S_2\right) = p_1 p_{2|\mathcal{J}_{-1}}.$$

Proof: The result follows immediately from the Markov property of order statistics (e.g. David (1981), Section 2.7). It can be derived from first principles as follows:

$$
\begin{aligned}
\pi &= \Pr\left(S_1 = \max \mathcal{S}, S_2 = \max \mathcal{S}_{-1}\right) \\
&= \Pr\left(S_1 > S_j, j \in \mathcal{J}_{-1}; S_2 > S_k, k \in \mathcal{J}_{-\{1,2\}}\right) \\
&= \Pr\left(S_1 > S_2; S_2 > S_k, k \in \mathcal{J}_{-\{1,2\}}\right) \\
&= \mathbb{E}_{\epsilon_2}\left[\Pr\left(S_1 > S_2 | S_2\right) \prod_{k>2} \Pr\left(S_k < S_2 | S_2\right)\right] \\
&= \mathbb{E}_{\epsilon_2}\left[\prod_{k>2} \exp\left(-\exp\left(-\left(\frac{\delta_2 - \delta_k}{\sigma} + \epsilon_2\right)\right)\right)\right. \\
&\quad \left.\left(1 - \exp\left(-\exp\left(-\left(\frac{\delta_2 - \delta_1}{\sigma} + \epsilon_2\right)\right)\right)\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \int \exp\left(-\epsilon_2 - \exp\left(-\epsilon_2\right)\left(1 + \sum_{k>2} \exp\left(\frac{\delta_k - \delta_2}{\sigma}\right)\right)\right) d\epsilon_2 \\
&\quad - \int \exp\left(-\epsilon_2 - \exp\left(-\epsilon_2\right)\left(1 + \sum_{k\neq 2} \exp\left(\frac{\delta_k - \delta_2}{\sigma}\right)\right)\right) d\epsilon_2 \\
&= \int \exp\left(-\epsilon_2 - \exp\left(-\epsilon_2 - \ln p_{2|\mathcal{J}_{-1}}\right)\right) d\epsilon_2 - \int \exp\left(-\epsilon_2 - \exp\left(-\epsilon_2 - \ln p_2\right)\right) d\epsilon_2 \\
&= p_{2|\mathcal{J}_{-1}} - p_2 \\
&= p_{2|\mathcal{J}_{-1}} p_1.
\end{aligned}
$$

□

Let $\mathcal{S}_{-12} = \{S_k, k > 2\}$, $I_{-12} = \sigma \ln\left(\sum_{k>2} \exp\left(\frac{\delta_k}{\sigma}\right)\right)$, and $p_{2|12} = \frac{\exp\left(\frac{\delta_2}{\sigma}\right)}{\exp\left(\frac{\delta_1}{\sigma}\right)+\exp\left(\frac{\delta_2}{\sigma}\right)}$.

Result 4: *Conditional distribution of the third highest surplus, given the first and second highest.* For $y \in \mathbb{R}$,

$$
\begin{aligned}
F_{S_{J-2:J}|S_{J-1:J},S_{J:J}}(y) &= \Pr\left(\max \mathcal{S}_{-\{1,2\}} < y | S_2 = \max \mathcal{S}_{-1}, S_1 = \max \mathcal{S}\right) \\
&= -\frac{1-p_1}{p_1} \exp\left(-\exp\left(-\frac{y - I}{\sigma}\right)\right) + \frac{1}{p_1} \exp\left(-\exp\left(-\frac{y - I_{-1}}{\sigma}\right)\right) \\
&\quad -\frac{1}{p_1}\frac{1-p_1}{p_1+p_2}\left[\exp\left(-\exp\left(-\frac{y - I_{-\{1,2\}}}{\sigma}\right)\right) - \exp\left(-\exp\left(-\frac{y - I}{\sigma}\right)\right)\right] \\
&\quad +\frac{1}{p_1 p_{2|\mathcal{J}_{-1}}}\left[\exp\left(-\exp\left(-\frac{y - I_{-\{1,2\}}}{\sigma}\right)\right) - \exp\left(-\exp\left(-\frac{y - I_{-1}}{\sigma}\right)\right)\right].
\end{aligned}
$$

Proof:

$$\Pr\left(\max\mathcal{S}_{-\{1,2\}} < y | S_2 = \max\mathcal{S}_{-1}, S_1 = \max\mathcal{S}\right) = \frac{\Pr\left(S_j < y, j > 2; S_2 = \max\mathcal{S}_{-1}, S_1 = \max\mathcal{S}\right)}{p_1 p_{2|\mathcal{J}_{-1}}}.$$

Consider the numerator:

$$
\begin{aligned}
P(y) &= \Pr\left(S_j < y, S_j < S_2, j > 2; S_k < S_1, k > 1\right) \\
&= \Pr\left(\epsilon_j < \frac{y-\delta_j}{\sigma}, \epsilon_j < \frac{\delta_2-\delta_j}{\sigma} + \epsilon_2, j > 2; \epsilon_2 < \frac{\delta_1-\delta_2}{\sigma} + \epsilon_1\right) \\
&= \Pr\left(\epsilon_j < \frac{\delta_2-\delta_j}{\sigma} + \min\left\{\frac{y-\delta_2}{\sigma}, \epsilon_2\right\}, j > 2;\right) \\
&= \mathbb{E}_{\epsilon_2}\left[\prod_{j>2}\left[\Pr\left(\epsilon_j < \frac{\delta_2-\delta_j}{\sigma} + \min\left\{\frac{y-\delta_2}{\sigma}, \epsilon_2\right\}\middle| \epsilon_2\right)\right]\left(1 - \Pr\left(\epsilon_1 < \frac{\delta_2-\delta_1}{\sigma} + \epsilon_2\middle| \epsilon_2\right)\right)\right] \\
&= \int_{-\infty}^{\frac{y-\delta_2}{\sigma}} \exp\left(-\sum_{j>2}\exp\left(-\left(\frac{\delta_2-\delta_j}{\sigma} + \epsilon_2\right)\right)\right)\left(1 - \exp\left(-\exp\left(-\left(\frac{\delta_2-\delta_1}{\sigma} + \epsilon_2\right)\right)\right)\right) \\
&\quad \exp\left(-\epsilon_2 - \exp\left(-\epsilon_2\right)\right) d\epsilon_2 \\
&\quad + \int_{\frac{y-\delta_2}{\sigma}}^{+\infty} \exp\left(-\sum_{j>2}\exp\left(-\frac{y-\delta_j}{\sigma}\right)\right)\left(1 - \exp\left(-\exp\left(-\left(\frac{\delta_2-\delta_1}{\sigma} + \epsilon_2\right)\right)\right)\right) \\
&\quad \exp\left(-\epsilon_2 - \exp(-\epsilon_2)\right) d\epsilon_2 \\
&= \int_{-\infty}^{\frac{y-\delta_2}{\sigma}} \exp\left(-\exp(-\epsilon_2)\sum_{j>2}\exp\left(\frac{\delta_j-\delta_2}{\sigma}\right)\right)\left(1 - \exp\left(-\exp(-\epsilon_2)\exp\left(\frac{\delta_1-\delta_2}{\sigma}\right)\right)\right) \\
&\quad \exp\left(-\epsilon_2 - \exp(-\epsilon_2)\right) d\epsilon_2 \\
&\quad \int_{\frac{y-\delta_2}{\sigma}}^{+\infty} \exp\left(-\exp\left(-\frac{y}{\sigma}\right)\sum_{j>2}\exp\left(\frac{\delta_j}{\sigma}\right)\right)\left(1 - \exp\left(-\exp(-\epsilon_2)\exp\left(\frac{\delta_1-\delta_2}{\sigma}\right)\right)\right) \\
&\quad \exp\left(-\epsilon_2 - \exp(-\epsilon_2)\right) d\epsilon_2 \\
&= -\int_{-\infty}^{\frac{y-\delta_2}{\sigma}} \exp\left(-\epsilon_2 - \exp(-\epsilon_2)\left(1 + \sum_{j\in\mathcal{J}_{-2}}\exp\left(\frac{\delta_j-\delta_2}{\sigma}\right)\right)\right) d\epsilon_2 \\
&\quad + \int_{-\infty}^{\frac{y-\delta_2}{\sigma}} \exp\left(-\epsilon_2 - \exp(-\epsilon_2)\left(1 + \sum_{j>2}\exp\left(\frac{\delta_j-\delta_2}{\sigma}\right)\right)\right) d\epsilon_2 \\
&\quad - \exp\left(-\exp\left(-\frac{y}{\sigma}\right)\sum_{j>2}\exp\left(\frac{\delta_j}{\sigma}\right)\right)\int_{\frac{y-\delta_2}{\sigma}}^{+\infty} \exp\left(-\epsilon_2 - \exp(-\epsilon_2)\left(1 + \exp\left(\frac{\delta_1-\delta_2}{\sigma}\right)\right)\right) d\epsilon_2 \\
&\quad + \exp\left(-\exp\left(-\frac{y - I_{-\{1,2\}}}{\sigma}\right)\right)\left(1 - \exp\left(-\exp\left(-\frac{y-\delta_2}{\sigma}\right)\right)\right) \\
&= \int_{-\infty}^{\frac{y-\delta_2}{\sigma}} \left[-\exp\left(-\epsilon_2 - \exp(-\epsilon_2 - \ln p_2)\right) + \exp\left(-\epsilon_2 - \exp\left(-\epsilon_2 - \ln p_{2|\mathcal{J}_{-1}}\right)\right)\right] d\epsilon_2 \\
&\quad - \exp\left(-\exp\left(-\frac{y - I_{-\{1,2\}}}{\sigma}\right)\right) p_{2|12}\left(1 - \exp\left(-\exp\left(-\left(\frac{y-\delta_2}{\sigma} + \ln p_{2|12}\right)\right)\right)\right) \\
&\quad + \exp\left(-\exp\left(-\frac{y - I_{-\{1,2\}}}{\sigma}\right)\right)\left(1 - \exp\left(-\exp\left(-\frac{y-\delta_2}{\sigma}\right)\right)\right)
\end{aligned}
$$

$$\begin{aligned}
= & -p_2 \exp\left(-\exp\left(-\left(\frac{y-\delta_2}{\sigma}+\ln p_2\right)\right)\right) + p_{2|\mathcal{J}_{-1}} \exp\left(-\exp\left(-\left(\frac{y-\delta_2}{\sigma}+\ln p_{2|\mathcal{J}_{-1}}\right)\right)\right) \\
& -\exp\left(-\exp\left(-\frac{y-I_{-\{1,2\}}}{\sigma}\right)\right) p_{2|12}\left(1-\exp\left(-\exp\left(-\left(\frac{y-\delta_2}{\sigma}+\ln p_{2|12}\right)\right)\right)\right) \\
& +\exp\left(\exp\left(-\frac{y-I_{-\{1,2\}}}{\sigma}\right)\right)\left(1-\exp\left(-\exp\left(-\frac{y-\delta_2}{\sigma}\right)\right)\right) \\
= & -p_2 \exp\left(-\exp\left(-\left(\frac{y-\delta_2}{\sigma}+\ln p_2\right)\right)\right) + p_{2|\mathcal{J}_{-1}} \exp\left(-\exp\left(-\left(\frac{y-\delta_2}{\sigma}+\ln p_{2|\mathcal{J}_{-1}}\right)\right)\right) \\
& -p_{2|12}\exp\left(-\exp\left(-\frac{y-I_{-\{1,2\}}}{\sigma}\right)\right) \\
& +p_{2|12}\exp\left(-\exp\left(-\frac{y}{\sigma}\right)\left(\exp\left(\frac{I_{-\{1,2\}}}{\sigma}\right)+\exp\left(\frac{\delta_2}{\sigma}-\ln p_{2|\mathcal{J}_{-1}}\right)\right)\right) \\
& +\exp\left(\exp\left(-\frac{y-I_{-\{1,2\}}}{\sigma}\right)\right)-\exp\left(-\exp\left(-\frac{y}{\sigma}\right)\exp\left(\frac{I_{-\{1,2\}}}{\sigma}+\exp\left(\frac{\delta_2}{\sigma}\right)\right)\right) \\
= & -p_2 \exp\left(-\exp\left(-\frac{y-I}{\sigma}\right)\right) + p_{2|\mathcal{J}_{-1}} \exp\left(-\exp\left(-\frac{y-I_{-1}}{\sigma}\right)\right) \\
& -p_{2|12}\exp\left(-\exp\left(-\frac{y-I_{-\{1,2\}}}{\sigma}\right)\right)+p_{2|12}\exp\left(-\exp\left(-\frac{y-I}{\sigma}\right)\right) \\
& +\exp\left(\exp\left(-\frac{y-I_{-\{1,2\}}}{\sigma}\right)\right)-\exp\left(-\exp\left(-\frac{y-I_{-1}}{\sigma}\right)\right).
\end{aligned}$$

Consistency check: $\lim_{y\to+\infty} P(y) = -p_2 + p_{2|\mathcal{J}_{-1}} = p_{2|\mathcal{J}_{-1}}(1-p_{1^c}) = p_{2|\mathcal{J}_{-1}}p_1$, where $p_{1^c} = \exp\left(\frac{I_{-1}}{\sigma}\right)/\exp\left(\frac{I}{\sigma}\right)$.

Furthermore, denoting $\pi = p_1 p_{2|\mathcal{J}_{-1}}$,

$$\begin{aligned}
p_2/\pi &= \frac{p_2}{p_1 p_{2|\mathcal{J}_{-1}}} = \frac{\sum_{j>1}\exp\left(\frac{\delta_j}{\sigma}\right)}{\exp\left(\frac{\delta_1}{\sigma}\right)} = \frac{1-p_1}{p_1} \\
\frac{p_{2|\mathcal{J}_{-1}}}{\pi} &= \frac{1}{p_1} \\
\frac{p_{2|12}}{\pi} &= \frac{\exp\left(\frac{\delta_2}{\sigma}\right)}{\exp\left(\frac{\delta_1}{\sigma}\right)+\exp\left(\frac{\delta_2}{\sigma}\right)}\frac{1}{p_1 p_{2|\mathcal{J}_{-1}}} \\
&= \frac{1}{\exp\left(\frac{\delta_1}{\sigma}\right)+\exp\left(\frac{\delta_2}{\sigma}\right)}\frac{\sum_{j\in\mathcal{J}}\exp\left(\frac{\delta_j}{\sigma}\right)}{\exp\left(\frac{\delta_1}{\sigma}\right)}\sum_{k\in\mathcal{J}_{-1}}\exp\left(\frac{\delta_k}{\sigma}\right) \\
&= \frac{1}{p_1}\frac{1-p_1}{p_1+p_2}.
\end{aligned}$$

□

Result 5: *Conditional mean of the third highest surplus, given the first and second highest.*

$$\begin{aligned}\mathbb{E}\left[\max \mathcal{S}_{-\{1,2\}} | S_1 = \max \mathcal{S}, S_2 = \max \mathcal{S}_{-1}\right] &= \alpha - \frac{1-p_1}{p_1} I + \frac{1}{p_1} I_{-1} - \frac{1-p_1}{p_1} \frac{1}{p_1+p_2} \sigma \ln(1-p_1-p_2) \\ &\quad + \frac{1}{p_1 p_{2|\mathcal{J}_{-1}}} \sigma \ln\left(1 - p_{2|\mathcal{J}_{-1}}\right)\end{aligned}$$

where α is a constant.

Proof: This follows from Result 4 and

$$\begin{aligned} I_{-\{1,2\}} - I &= \sigma \ln \left(\frac{\sum_{k>2} \exp\left(\frac{\delta_k}{\sigma}\right)}{\sum_{j \in \mathcal{J}} \exp\left(\frac{\delta_j}{\sigma}\right)} \right) \\ &= \sigma \ln(1 - p_1 - p_2) \\ I_{-\{1,2\}} - I_{-1} &= \sigma \ln \left(\frac{\sum_{k>2} \exp\left(\frac{\delta_k}{\sigma}\right)}{\sum_{j \in \mathcal{J}_{-1}} \exp\left(\frac{\delta_j}{\sigma}\right)} \right) \\ &= \sigma \ln\left(1 - p_{2|\mathcal{J}_{-1}}\right). \end{aligned}$$

□

Result 6: *Probability of first, second and third highest surplus.*

$$\Pr\left(S_1 = \max \mathcal{S}, S_2 = \max \mathcal{S}_{-1}, S_3 = \max \mathcal{S}_{-\{1,2\}}\right) = p_1 p_{2|\mathcal{J}_{-1}} p_{3|\mathcal{J}_{-\{1,2\}}}.$$

Proof 1: This follows from the Markov property of order statistics, and from Result 3, by induction. □

Proof from first principles: Let

$$\begin{aligned} \pi_2 &= \Pr\left(S_1 = \max \mathcal{S}, S_2 = \max \mathcal{S}_{-1}, S_3 = \max \mathcal{S}_{-\{1,2\}}\right) \\ &= \Pr\left(S_1 > S_2, S_2 > S_3; S_3 > S_k, k > 3\right) \\ &= \Pr\left(\epsilon_2 < \frac{\delta_1 - \delta_2}{\sigma} + \epsilon_1, \epsilon_3 < \frac{\delta_2 - \delta_3}{\sigma} + \epsilon_2; \epsilon_k < \frac{\delta_3 - \delta_k}{\sigma} + \epsilon_3, k > 3\right) \\ &= \mathbb{E}_{\epsilon_3}\left[\prod_{k>3} \Pr\left(\epsilon_k < \frac{\delta_3 - \delta_k}{\sigma} + \epsilon_3 | \epsilon_3\right) \Pr\left(\epsilon_2 < \frac{\delta_1 - \delta_2}{\sigma} + \epsilon_1, \frac{\delta_3 - \delta_2}{\sigma} + \epsilon_3 < \epsilon_2 \middle| \epsilon_3\right)\right] \\ &= \mathbb{E}_{\epsilon_3}\left[\prod_{k>3} \Pr\left(\epsilon_k < \frac{\delta_3 - \delta_k}{\sigma} + \epsilon_3 \middle| \epsilon_3\right)\right. \\ &\qquad \left.\mathbb{E}_{\epsilon_1}\left[\Pr\left(\frac{\delta_3 - \delta_2}{\sigma} + \epsilon_3 < \epsilon_2 < \frac{\delta_1 - \delta_2}{\sigma} + \epsilon_1 \middle| \epsilon_1, \epsilon_3\right) \middle| \epsilon_1 > \frac{\delta_3 - \delta_1}{\sigma} + \epsilon_3\right]\right] \end{aligned}$$

Consider the inner conditional expectation, and define $p_{1|12} = p_{1|\mathcal{J}_{12}} = \frac{\exp\left(\frac{\delta_1}{\sigma}\right)}{\exp\left(\frac{\delta_1}{\sigma}\right)+\exp\left(\frac{\delta_2}{\sigma}\right)}$ and similarly for $p_{3|\mathcal{J}_{-\{1,2\}}}$:

$$
\begin{aligned}
M &= \mathbb{E}_{\epsilon_1}\left[\Pr\left(\frac{\delta_3-\delta_2}{\sigma}+\epsilon_3<\epsilon_2<\frac{\delta_1-\delta_2}{\sigma}+\epsilon_1\middle|\epsilon_1,\epsilon_3\right)\middle|\epsilon_1>\frac{\delta_3-\delta_1}{\sigma}+\epsilon_3\right]\\
&= \int_{\frac{\delta_3-\delta_1}{\sigma}+\epsilon_3}^{+\infty}\left[\exp\left(-\exp\left(-\left(\frac{\delta_1-\delta_2}{\sigma}+\epsilon_1\right)\right)\right)-\exp\left(-\exp\left(-\left(\frac{\delta_3-\delta_2}{\sigma}+\epsilon_3\right)\right)\right)\right]\\
&\quad \exp(-\epsilon_1-\exp(-\epsilon_1))d\epsilon_1\\
&= \int_{\frac{\delta_3-\delta_1}{\sigma}+\epsilon_3}^{+\infty}\exp\left(-\epsilon_1-\exp(-\epsilon_1)\left(1+\exp\left(\frac{\delta_2-\delta_1}{\sigma}\right)\right)\right)d\epsilon_1\\
&\quad -\exp\left(-\exp\left(-\left(\frac{\delta_3-\delta_2}{\sigma}+\epsilon_3\right)\right)\right)\left(1-\exp\left(-\exp\left(-\left(\frac{\delta_3-\delta_1}{\sigma}+\epsilon_3\right)\right)\right)\right)\\
&= p_{1|12}\left(1-\exp\left(-\exp\left(-\left(\frac{\delta_3-\delta_1}{\sigma}+\epsilon_3+\ln p_{1|12}\right)\right)\right)\right)\\
&\quad -\exp\left(-\exp\left(-\left(\frac{\delta_3-\delta_2}{\sigma}+\epsilon_3\right)\right)\right)+\exp\left(-\exp\left(-\epsilon_3\right)\left(\exp\left(\frac{\delta_2-\delta_3}{\sigma}\right)+\exp\left(\frac{\delta_1-\delta_3}{\sigma}\right)\right)\right)
\end{aligned}
$$

Inserting back into the outer expectation,

$$
\begin{aligned}
\pi_2 &= \mathbb{E}_{\epsilon_3}\left[\prod_{k>3}\exp\left(-\exp\left(-\left(\frac{\delta_3-\delta_k}{\sigma}\right)\right)\right)M\right]\\
&= p_{1|12}p_{3|\mathcal{J}_{-\{1,2\}}}-p_{1|12}p_3-p_{3|\mathcal{J}_{-1}}+p_3\\
&= p_{1|12}\left[p_{3|\mathcal{J}_{-\{1,2\}}}-p_3\right]-p_{3|\mathcal{J}_{-1}}p_1\\
&= p_{1|12}p_{3|\mathcal{J}_{-\{1,2\}}}p_{12}-p_{3|\mathcal{J}_{-1}}p_1\\
&= p_1\left[p_{3|\mathcal{J}_{-\{1,2\}}}-p_{3|\mathcal{J}_{-1}}\right]\\
&= p_1p_{2|\mathcal{J}_{-1}}p_{3|\mathcal{J}_{-\{1,2\}}}.
\end{aligned}
$$

□

Result 7: *Conditional distribution of* $\max\mathcal{S}_{-\{1,2,3\}}$*, given* $S_1 > S_2 > S_3 > \max\mathcal{S}_{-\{1,2,3\}}$.

$$\Pr\left(\max\mathcal{S}_{-\{1,2,3\}}<y|S_1=\max\mathcal{S},S_2=\max\mathcal{S}_{-1},S_3=\max\mathcal{S}_{-\{1,2\}}\right)=\frac{P(y)}{R},$$

where R is as in Result 8, and

$$\begin{aligned}
P(y) &= A+B+C+D+E+F+G+H \\
A &= p_{1|12}p_{3|\mathcal{J}_{-\{1,2\}}}\exp\left(-\exp\left(-\frac{y-I_{-\{1,2\}}}{\sigma}\right)\right) \\
B &= -p_{1|12}p_3\exp\left(-\exp\left(-\frac{y-I}{\sigma}\right)\right) \\
C &= -p_{3|\mathcal{J}_{-1}}\exp\left(-\exp\left(-\frac{y-I_{-1}}{\sigma}\right)\right) \\
D &= p_3\exp\left(-\exp\left(-\frac{y-I}{\sigma}\right)\right) \\
E &= p_{1|12}\exp\left(-\exp\left(-\frac{y-I_{-\{1,2,3\}}}{\sigma}\right)\right)\left(1-\exp\left(-\exp\left(-\frac{y-\delta_3}{\sigma}\right)\right)\right) \\
F &= -p_{1|12}\exp\left(-\exp\left(-\frac{y-I_{-\{1,2,3\}}}{\sigma}\right)\right)p_{3|123}\left(1-\exp\left(-\exp\left(-\frac{y-I_{123}}{\sigma}\right)\right)\right) \\
G &= -\exp\left(-\exp\left(-\frac{y-I_{-\{1,2,3\}}}{\sigma}\right)\right)p_{3|23}\left(1-\exp\left(-\exp\left(-\frac{y-I_{23}}{\sigma}\right)\right)\right) \\
H &= \exp\left(-\exp\left(-\frac{y-I_{-\{1,2,3\}}}{\sigma}\right)\right)p_{3|123}\left(1-\exp\left(-\exp\left(-\frac{y-I_{123}}{\sigma}\right)\right)\right).
\end{aligned}$$

Proof: R follows from Result 6. Furthermore, for any $y \in \mathbb{R}$,

$$\begin{aligned}
P(y) &= \Pr\left(\max\mathcal{S}_{-\{1,2,3\}} < y, S_1 > S_2 > S_3 > \max\mathcal{S}_{-\{1,2,3\}}\right) \\
&= \Pr\left(\{\epsilon_j < \frac{\delta_3-\delta_j}{\sigma} + \min\{\frac{y-\delta_3}{\sigma}, \epsilon_3\}, j \in \{4,\cdots,J\}\}, \epsilon_3 + \frac{\delta_3-\delta_2}{\sigma} < \epsilon_2 < \epsilon_1 + \frac{\delta_1-\delta_2}{\sigma}\right) \\
&= \mathbb{E}_{\epsilon_3}\left[\prod_{j\geq 4}\Pr\left(\epsilon_j < \frac{\delta_3-\delta_j}{\sigma} + \min\{\frac{y-\delta_3}{\sigma}, \epsilon_3\}\middle|\epsilon_3\right)\Pr\left(\epsilon_3 + \frac{\delta_3-\delta_2}{\sigma} < \epsilon_2 < \epsilon_1 + \frac{\delta_1-\delta_2}{\sigma}\middle|\epsilon_3\right)\right] \\
&= \mathbb{E}_{\epsilon_3}\left[\prod_{j\geq 4}\Pr\left(\epsilon_j < \frac{\delta_3-\delta_j}{\sigma} + \min\{\frac{y-\delta_3}{\sigma}, \epsilon_3\}\middle|\epsilon_3\right)\right. \\
&\qquad \left.\mathbb{E}_{\epsilon_1}\left[\Pr\left(\epsilon_3 + \frac{\delta_3-\delta_2}{\sigma} < \epsilon_2 < \epsilon_1 + \frac{\delta_1-\delta_2}{\sigma}\middle|\epsilon_1,\epsilon_3\right); \epsilon_1 > \frac{\delta_3-\delta_1}{\sigma} + \epsilon_3\right]\right] \\
&= \mathbb{E}_{\epsilon_3}\left[\prod_{j\geq 4}\Pr\left(\epsilon_j < \frac{\delta_3-\delta_j}{\sigma} + \min\{\frac{y-\delta_3}{\sigma}, \epsilon_3\}\middle|\epsilon_3\right)M\right],
\end{aligned}$$

where M is as defined in the proof of Result 6. Therefore,

$$\begin{aligned}
P(y) &= \mathbb{E}_{\epsilon_3}\Bigg[\Bigg[\exp\left(-\exp\left(-\epsilon_3)(\sum_{j\geq 4}\exp\left(\frac{\delta_j-\delta_3}{\sigma}\right)\right)\right)1_{\{\epsilon_3\leq\frac{y-\delta_3}{\sigma}\}} \\
&\exp\left(-\exp\left(-\frac{y}{\sigma}\right)\left(\sum_{j\geq 4}\exp\left(\frac{\delta_j}{\sigma}\right)\right)\right)1_{\{\epsilon_3>\frac{y-\delta_3}{\sigma}\}}\Bigg] \\
&\Bigg[p_{1|12}\left(1-\exp\left(-\exp\left(-\epsilon_3\right)\left(\exp\left(\frac{\delta_1-\delta_3}{\sigma}\right)+\exp\left(\frac{\delta_2-\delta_3}{\sigma}\right)\right)\right)\right) \\
&-\exp\left(-\exp\left(-\epsilon_3\right)\exp\left(\frac{\delta_2-\delta_3}{\sigma}\right)\right) \\
&+\exp\left(-\exp\left(-\epsilon_3\right)\left(\exp\left(\frac{\delta_1-\delta_3}{\sigma}\right)+\exp\left(\frac{\delta_2-\delta_3}{\sigma}\right)\right)\right)\Bigg]\Bigg]
\end{aligned}$$

In this expression,

$$\begin{aligned}
A &= \int_{-\infty}^{\frac{y-\delta_3}{\sigma}}\exp\left(-\exp(-\epsilon_3)\left(\sum_{j\geq 4}\exp\left(\frac{\delta_j-\delta_3}{\sigma}\right)\right)\right)p_{1|12}\exp(-\epsilon_3-\exp(-\epsilon_3))d\epsilon_3 \\
&= p_{1|12}p_{3|\mathcal{J}_{-\{1,2\}}}\exp\left(-\exp\left(-\frac{y-\delta_3}{\sigma}-\ln p_{3|\mathcal{J}_{-\{1,2\}}}\right)\right) \\
&= p_{1|12}p_{3|\mathcal{J}_{-\{1,2\}}}\exp\left(-\exp\left(-\frac{y-I_{-\{1,2\}}}{\sigma}\right)\right) \\
B &= -\int_{-\infty}^{\frac{y-\delta_3}{\sigma}}\exp\left(-\exp(-\epsilon_3)\left(\sum_{j\geq 4}\exp\left(\frac{\delta_j-\delta_3}{\sigma}\right)\right)\right)p_{1|12}\times \\
&\exp\left(-\exp(-\epsilon_3)\left(\exp\left(\frac{\delta_1-\delta_3}{\sigma}\right)+\exp\left(\frac{\delta_2-\delta_3}{\sigma}\right)\right)\right)\exp(-\epsilon_3-\exp(-\epsilon_3))d\epsilon_3 \\
&= -p_{1|12}p_3\exp\left(-\exp\left(-\frac{y-\delta_3}{\sigma}-\ln p_3\right)\right) \\
&= -p_{1|12}p_3\exp\left(-\exp\left(-\frac{y-I}{\sigma}\right)\right) \\
C &= -\int_{-\infty}^{\frac{y-\delta_3}{\sigma}}\exp\left(-\exp(-\epsilon_3)\left(\sum_{j\geq 4}\exp\left(\frac{\delta_j-\delta_3}{\sigma}\right)\right)\right)\times \\
&\exp\left(-\exp(-\epsilon_3)\exp\left(\frac{\delta_2-\delta_3}{\sigma}\right)\right)\exp(-\epsilon_3-\exp(-\epsilon_3))d\epsilon_3 \\
&= -p_{3|\mathcal{J}_{-1}}\exp\left(-\exp\left(-\frac{y-I_{-1}}{\sigma}\right)\right)
\end{aligned}$$

$$
\begin{aligned}
D &= \int_{-\infty}^{\frac{y-\delta_3}{\sigma}} \exp\left(-\exp(-\epsilon_3)\left(\sum_{j\geq 4}\exp\left(\frac{\delta_j-\delta_3}{\sigma}\right)\right)\right) \times \\
&\quad \exp\left(-\exp(-\epsilon_3)\left(\exp\left(\frac{\delta_1-\delta_3}{\sigma}\right)+\exp\left(\frac{\delta_2-\delta_3}{\sigma}\right)\right)\right)\exp(-\epsilon_3-\exp(-\epsilon_3))d\epsilon_3 \\
&= p_3\exp\left(-\exp\left(-\frac{y-I}{\sigma}\right)\right) \\
E &= \int_{\frac{y-\delta_3}{\sigma}}^{\infty} \exp\left(-\exp\left(-\frac{y}{\sigma}\right)\left(\sum_{j\geq 4}\exp\left(\frac{\delta_j}{\sigma}\right)\right)\right) p_{1|12}\exp(-\epsilon_3-\exp(-\epsilon_3))d\epsilon_3 \\
&= p_{1|12}\exp\left(-\exp\left(-\frac{y-I_{-\{1,2,3\}}}{\sigma}\right)\right)\left(1-\exp\left(-\exp\left(-\frac{y-\delta_3}{\sigma}\right)\right)\right) \\
F &= -\int_{\frac{y-\delta_3}{\sigma}}^{\infty} \exp\left(-\exp\left(-\frac{y}{\sigma}\right)\left(\sum_{j\geq 4}\exp\left(\frac{\delta_j}{\sigma}\right)\right)\right) \times \\
&\quad p_{1|12}\exp\left(-\exp(-\epsilon_3)\left(\exp\left(\frac{\delta_1-\delta_3}{\sigma}\right)+\exp\left(\frac{\delta_2-\delta_3}{\sigma}\right)\right)\right)\exp(-\epsilon_3-\exp(-\epsilon_3))d\epsilon_3 \\
&= -p_{1|12}\exp\left(-\exp\left(-\frac{y-I_{-\{1,2,3\}}}{\sigma}\right)\right)p_{3|123}\left(1-\exp\left(-\exp\left(-\frac{y-I_{123}}{\sigma}\right)\right)\right) \\
G &= -\int_{\frac{y-\delta_3}{\sigma}}^{\infty} \exp\left(-\exp\left(-\frac{y}{\sigma}\right)\left(\sum_{j\geq 4}\exp\left(\frac{\delta_j}{\sigma}\right)\right)\right)\exp\left(-\exp(-\epsilon_3)\exp\left(\frac{\delta_2-\delta_3}{\sigma}\right)\right) \times \\
&\quad \exp(-\epsilon_3-\exp(-\epsilon_3))d\epsilon_3 \\
&= -\exp\left(-\exp\left(-\frac{y-I_{-\{1,2,3\}}}{\sigma}\right)\right)p_{3|23}\left(1-\exp\left(-\exp\left(-\frac{y-I_{23}}{\sigma}\right)\right)\right) \\
H &= \int_{\frac{y-\delta_3}{\sigma}}^{\infty} \exp\left(-\exp\left(-\frac{y}{\sigma}\right)\left(\sum_{j\geq 4}\exp\left(\frac{\delta_j}{\sigma}\right)\right)\right) \\
&\quad \exp\left(-\exp(-\epsilon_3)\left(\exp\left(\frac{\delta_1-\delta_3}{\sigma}\right)+\exp\left(\frac{\delta_2-\delta_3}{\sigma}\right)\right)\right)\exp(-\epsilon_3-\exp(-\epsilon_3))d\epsilon_3 \\
&= \exp\left(-\exp\left(-\frac{y-I_{-\{1,2,3\}}}{\sigma}\right)\right)p_{3|123}\left(1-\exp\left(-\exp\left(-\frac{y-I_{123}}{\sigma}\right)\right)\right),
\end{aligned}
$$

from which the result follows. □

Consistency check:

$$
\lim_{y\to+\infty} P(y) = p_{1|12}p_{3|\mathcal{J}_{-\{1,2\}}} - p_{1|12}p_3 - p_{3|\mathcal{J}_{-1}}p_1 = p_1p_{2|\mathcal{J}_{-1}}p_{3|\mathcal{J}_{-\{1,2\}}},
$$

□

Result 8: *Conditional mean of* $\max \mathcal{S}_{-\{1,2,3\}}$, *given* $S_1 = \max \mathcal{S}, S_2 = \max \mathcal{S}_{-1}, S_3 = \max \mathcal{S}_{-\{1,2\}}$.

$$\mathbb{E}\left[\max \mathcal{S}_{-\{1,2,3\}} | S_1 = \max \mathcal{S}, S_2 = \max \mathcal{S}_{-1}, S_3 = \max \mathcal{S}_{-\{1,2\}}\right] = \beta + \frac{Q}{R},$$

where β is a constant, R is as in Result 8 and

$$\begin{aligned} Q &= p_{1|12} p_{3|\mathcal{J}_{-\{1,2\}}} I_{-\{1,2\}} - p_{1|12} p_3 I - p_{3|\mathcal{J}_{-1}} I_{-1} + p_3 I + p_{1|12} \frac{1}{\sigma} \left(I_{-\{1,2,3\}} - I_{-\{1,2\}} \right) \\ & \quad - p_{1|12} p_{3|123} \frac{1}{\sigma} \left(I_{-\{1,2,3\}} - I \right) - p_{3|23} \frac{1}{\sigma} \left(I_{-\{1,2,3\}} - I_{-1} \right) + p_{3|123} \frac{1}{\sigma} \left(I_{-\{1,2,3\}} - I \right). \end{aligned}$$

Proof: The result follows from Result 7, and (except for constants that involve Euler's constant and are subsumed in β)

$$\begin{aligned} \int A dy &= p_{1|12} p_{3|\mathcal{J}_{-\{1,2\}}} I_{-\{1,2\}} \\ \int B dy &= -p_{1|12} p_3 I \\ \int C dy &= -p_{3|\mathcal{J}_{-1}} I_{-1} \\ \int D dy &= p_3 I \\ \int E dy &= p_{1|12} \left[\frac{I_{-\{1,2,3\}}}{\sigma} - \frac{I_{-\{1,2\}}}{\sigma} \right] \\ \int F dy &= -p_{1|12} p_{3|\mathcal{J}_{1,2,3}} \left[\frac{I_{-\{1,2,3\}}}{\sigma} - \frac{I}{\sigma} \right] \\ \int G dy &= -p_{3|23} \left[\frac{I_{-\{1,2,3\}}}{\sigma} - \frac{I_{-1}}{\sigma} \right] \\ \int H dy &= p_{3|\mathcal{J}_{123}} \left[\frac{I_{-\{1,2,3\}}}{\sigma} - \frac{I}{\sigma} \right]. \end{aligned}$$

□

4 Extensions to the Generalized Extreme Value Distribution

4.1 Analytical Results

The generalized extreme value (GEV) distribution is the basis of the nested multinomial logit (NMNL) model. Consider the specific case in which a decision maker has one outside option and $J = \#\mathcal{J}$ inside options; the GEV CDF of this model, henceforth referred to as model G, is

$$F\left(\epsilon_j, j \in \{0, \mathcal{J}\}\right) = \exp\left(-\exp\left(-\frac{\epsilon_0}{\sigma}\right) - \left(\sum_{j\in\mathcal{J}} \exp\left(-\frac{\epsilon_j}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}\right),$$

where $\lambda \in [0,1]$ captures the correlation of the inside options. When a researcher analyzes a consumer's choice from a particular group of products (e.g., automobiles), all other alternatives are lumped into the outside option. Thus, in applications, it is very important to allow the correlation between shocks of inside products and correlation between the outside option and an inside product to differ from each other.

Notice that

$$F(\epsilon_j, j \in \mathcal{J}_{-1}) = \exp\left(-\left(\sum_{j\in\mathcal{J}_{-1}} \exp\left(-\frac{\epsilon_j}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}\right),$$

and

$$F(\epsilon_j) = \exp\left(-\exp\left(-\frac{\epsilon_j}{\sigma}\right)\right), \ \ j = 0, 1, \cdots, J.$$

Also,

$$\begin{aligned}\frac{\partial}{\partial \epsilon_1} F(\epsilon_j, j \in \mathcal{J}) &= \frac{1}{\sigma}\exp\left(-\frac{\epsilon_1}{\sigma(1-\lambda)}\right)\left(\sum_{j\in\mathcal{J}} \exp\left(-\frac{\epsilon_j}{\sigma(1-\lambda)}\right)\right)^{-\lambda} \cdots \\ &\quad \times \exp\left(-\left(\sum_{j\in\mathcal{J}} \exp\left(-\frac{\epsilon_j}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}\right).\end{aligned}$$

Finally, let P_1 denote the NMNL choice probability of alternative 1,

$$\begin{aligned} P_1 &= \frac{\exp\left(\frac{\delta_1}{\sigma(1-\lambda)}\right)\left(\sum_{j\in\mathcal{J}}\exp\left(\frac{\delta_j}{\sigma(1-\lambda)}\right)\right)^{-\lambda}}{\exp\left(\frac{\delta_0}{\sigma}\right)+\left(\sum_{j\in\mathcal{J}}\exp\left(\frac{\delta_j}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}} \\ &= \frac{\left(\sum_{j\in\mathcal{J}}\exp\left(\frac{\delta_j-\delta_1}{\sigma(1-\lambda)}\right)\right)^{-\lambda}}{\exp\left(\frac{\delta_0-\delta_1}{\sigma}\right)+\left(\sum_{j\in\mathcal{J}}\exp\left(\frac{\delta_j-\delta_1}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}}, \end{aligned}$$

which reduces to the MNL choice probability p_1 when $\lambda = 0$.

In the MNL model with i.i.d. EV(0,σ) residuals, it is well known that

$$\begin{aligned} \mathbb{E}[S_1|S_1 = \max \mathcal{J}] &= \delta_1 - \sigma \ln p_1 + \sigma e \\ &= \delta_1 + \sigma \ln \left[1 + \sum_{j\in\mathcal{J}_{-1}} \exp\left(\frac{\delta_j - \delta_1}{\sigma}\right)\right] + \sigma e, \end{aligned}$$

where p_1 is the MNL choice probability for alternative 1 and e is Euler's constant.

The corresponding result for the NMNL model is

Result 9: *Under model G,*

$$\mathbb{E}\left[S_1|S_1 = \max\left\{\mathcal{J} \cup \{0\}\right\}\right] = \delta_1 + \sigma \ln D_1 + \sigma\gamma,$$

where $D_1 = \exp\left(\frac{\delta_0-\delta_1}{\sigma}\right) + \left(\sum_{j\in\mathcal{J}}\exp\left(\frac{\delta_j-\delta_1}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}$.

Proof: Start by considering

$$\Pr\left(S_1 < y|S_1 = \max\left\{\mathcal{J}\cup\{0\}\right\}\right) = \frac{\Pr\left(S_1 < y \ \& \ S_1 = \max\left\{\mathcal{J}\cup\{0\}\right\}\right)}{P_1}.$$

Then,

$$
\begin{aligned}
P(y) &= \Pr\left(S_1 < y \;\&\; S_1 = \max\{\mathcal{J} \cup \{0\}\}\right) \\
&= \int_{-\infty}^{y} \Pr(S_0 < s | S_1 = s) \Pr(S_j < s, j \in \mathcal{J}_{-1} | S_1 = s) f_{S_1}(s) ds \\
&= \int_{-\infty}^{y} \Pr(S_0 < s | S_1 = s) \Pr(S_j < s, j \in \mathcal{J}_{-1} \;\&\; S_1 = s) ds \\
&= \int_{-\infty}^{y-\delta_1} \exp\left(-\exp\left(-\frac{\epsilon_1}{\sigma}\right)\left[\exp\left(\frac{\delta_0 - \delta_1}{\sigma}\right) + \left(\sum_{j \in \mathcal{J}} \exp\left(\frac{\delta_j - \delta_1}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}\right]\right) \\
&\quad \left(\sum_{j \in \mathcal{J}} \exp\left(\frac{\delta_j - \delta_1}{\sigma(1-\lambda)}\right)\right)^{-\lambda} \frac{1}{\sigma} \exp\left(\frac{\epsilon_1}{\sigma}\right) d\epsilon_1 \\
&= \left(\sum_{j \in \mathcal{J}} \exp\left(\frac{\delta_j - \delta_1}{\sigma(1-\lambda)}\right)\right)^{-\lambda} \int_{-\infty}^{y-\delta_1} \frac{1}{\sigma} \exp\left(\frac{\epsilon_1}{\sigma}\right) \exp\left(-\exp\left(-\frac{\epsilon_1}{\sigma}\right) D_1\right) d\epsilon_1 \\
&= P_1 \int_{-\infty}^{y-\delta_1} \frac{1}{\sigma} \exp\left(-\frac{\epsilon_1 - \sigma \ln D_1}{\sigma} - \exp\left(-\frac{\epsilon_1 - \sigma \ln D_1}{\sigma}\right)\right) d\epsilon_1,
\end{aligned}
$$

from which the result follows. □

Note that Result 9 implies the well-known result for the conditional mean in the MNL model. Also, since

$$
D_1 = \exp\left(-\frac{\delta_1}{\sigma}\right)\left(\exp\left(\frac{\delta_0}{\sigma}\right) + \left(\sum_{j \in \mathcal{J}} \exp\left(\frac{\delta_j}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}\right),
$$

Result 9 implies the symmetry property,

$$
\begin{aligned}
\mathbb{E}\left[S_j | S_j = \max\{\mathcal{J} \cup \{0\}\}\right] &= \mathbb{E}\left[S_k | S_k = \max\{\mathcal{J} \cup \{0\}\}\right] \\
&= \sigma \ln \left(\exp\left(\frac{\delta_0}{\sigma}\right) + \left(\sum_{j \in \mathcal{J}} \exp\left(\frac{\delta_j}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}\right),
\end{aligned}
$$

for any $j, k \in \mathcal{J}$.

Result 10: The distribution of the maximum under model G is EV$(I(\lambda), \sigma)$, where

$$
I(\lambda) = \sigma \ln \left[\exp\left(\frac{\delta_0}{\sigma}\right) + \left(\sum_{j \in \mathcal{J}} \exp\left(\frac{\delta_j}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}\right].
$$

Proof: This follows from

$$
\begin{aligned}
P(y) &= \Pr\left(\max\{\mathcal{J} \cup \{0\}\} < y\right) \\
&= \Pr\left(S_j < y, j = 0, 1 \cdots, J\right) \\
&= \exp\left(-\exp\left(-\frac{y-\delta_0}{\sigma}\right) - \left(\sum_{j\in\mathcal{J}} \exp\left(-\frac{y-\delta_j}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}\right) \\
&= \exp\left(-\exp\left(\frac{y}{\sigma}\right)\left[\exp\left(\frac{\delta_0}{\sigma}\right) + \left(\sum_{j\in\mathcal{J}} \exp\left(\frac{\delta_j}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}\right]\right) \\
&= \exp\left(-\exp\left(-\frac{1}{\sigma}(y - \sigma I(\lambda))\right)\right).
\end{aligned}
$$

□

Notice that $I(0) = I$, the conventional inclusive value and expectation of the maximum in a standard logit model.

Let $M_{-1} = \max\{S_j, j \in \{\mathcal{J}_{-1} \cup \{0\}\}\}$. A corollary to this result is that the distribution of M_{-1} is $\mathrm{EV}(I_{-1}(\lambda), \sigma)$, where

$$
I_{-1}(\lambda) = \sigma \ln\left[\exp\left(\frac{\delta_0}{\sigma}\right) + \left(\sum_{j\in\mathcal{J}_{-1}} \exp\left(\frac{\delta_j}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}\right].
$$

Result 11: *The conditional expectation of the second-highest surplus, given the highest surplus, in model G.*

$$
\mathbb{E}\left[M_{-1} | S_1 = \max\{S_j, j \in \mathcal{J} \cup \{0\}\}\right] = \delta_1 + \sigma \ln D_1 + \sigma\gamma + \frac{1}{P_1}\left[I(\lambda)_{-1} - I(\lambda)\right],
$$

P_1 the NMNL choice probability of alternative 1.

Proof: Notice first that

$$
\begin{aligned}
P(y) &= \Pr\left(\max \mathcal{S}_{-1} < y | S_1 = \max\{S_j, j \in \mathcal{J} \cup \{0\}\}\right) \\
&= \Pr\left(\max \mathcal{S}_j < y, j \in \mathcal{J}_{-1} \ \& \ S_j < S_1, j \in \mathcal{J}_{-1} \cup \{0\}\right)/P_1,
\end{aligned}
$$

where P_1 is the NMNL choice probability for alternative 1.

Next, consider the numerator of the above expression,

$$
\begin{aligned}
N(y) &= \Pr\left(\max \mathcal{S}_j < y, j \in \mathcal{J}_{-1} \ \& \ S_j < S_1, j \in \mathcal{J}_{-1} \cup \{0\}\right) \\
&= \mathbb{E}_{\epsilon_1}\left[\Pr\left(\epsilon_j < \delta_1 - \delta_j + \epsilon_1, j \in \mathcal{J}_{-1} \cup \{0\} \middle| \epsilon_1 < y - \delta_1\right)\right] \\
&\quad + \mathbb{E}_{\epsilon_1}\left[\Pr\left(\epsilon_j < y - \delta_j, j \in \mathcal{J}_{-1} \cup \{0\} \middle| \epsilon_1 \geq y - \delta_1\right)\right] \\
&= \mathbb{E}_{\epsilon_1}\left[\Pr\left(\epsilon_j < \delta_1 - \delta_j + \epsilon_1, j \in \mathcal{J}_{-1} \middle| \epsilon_1 < y - \delta_1\right) \Pr\left(\epsilon_0 < \delta_1 - \delta_0 + \epsilon_1 \middle| \epsilon_1 < y - \delta_1\right)\right] \\
&\quad + \mathbb{E}_{\epsilon_1}\left[\Pr\left(\epsilon_j < y - \delta_j, j \in \mathcal{J}_{-1} \middle| \epsilon_1 \geq y - \delta_1\right) \Pr\left(\epsilon_0 < y - \delta_0 \middle| \epsilon_1 \geq y - \delta_1\right)\right] \\
&= \int_{-\infty}^{y-\delta_1} \Pr\left(\epsilon_j < \delta_1 - \delta_j + \epsilon_1, j \in \mathcal{J}_{-1}, \epsilon_1\right) \Pr\left(\epsilon_0 < \delta_1 - \delta_0 + \epsilon_1 | \epsilon_1\right) d\epsilon_1 \\
&\quad + \int_{y-\delta_1}^{\infty} \Pr\left(\epsilon_j < y - \delta_j, j \in \mathcal{J}_{-1}, \epsilon_1\right) \Pr\left(\epsilon_0 < y - \delta_0\right) d\epsilon_1 \\
&= \int_{-\infty}^{y-\delta_1} \frac{1}{\sigma} \exp\left(-\frac{\epsilon_1}{\sigma} - \exp\left(-\frac{\epsilon_1}{\sigma}\right)\left[\exp\left(\frac{\delta_0 - \delta_1}{\sigma}\right) + \left(\sum_{j \in \mathcal{J}} \exp\left(\frac{\delta_j - \delta_1}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}\right]\right) \\
&\quad \left(\sum_{j \in \mathcal{J}} \exp\left(\frac{\delta_j - \delta_1}{\sigma(1-\lambda)}\right)\right)^{-\lambda} d\epsilon_1 \\
&\quad + \int_{y-\delta_1}^{\infty} \frac{1}{\sigma} \exp\left(-\frac{\epsilon_1}{\sigma(1-\lambda)}\right) \exp\left(-\left(\exp\left(-\frac{\epsilon_1}{\sigma(1-\lambda)}\right) + \sum_{j \in \mathcal{J}_{-1}} \exp\left(-\frac{y - \delta_j}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}\right) \\
&\quad \left(\exp\left(-\frac{\epsilon_1}{\sigma(1-\lambda)}\right) + \sum_{j \in \mathcal{J}_{-1}} \exp\left(-\frac{y - \delta_j}{\sigma(1-\lambda)}\right)\right)^{-\lambda} \exp\left(-\exp\left(-\frac{y - \delta_0}{\sigma}\right)\right) d\epsilon_1 \\
&= P_1 \exp\left(-\exp\left(-\frac{y - \delta_1 - \sigma \ln D_1}{\sigma}\right)\right) \\
&\quad + \exp\left(-\exp\left(-\frac{y}{\sigma}\right)\left(\exp\left(\frac{\delta_0}{\sigma}\right) + \left(\sum_{j \in \mathcal{J}_{-1}} \exp\left(\frac{\delta_j}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}\right)\right) \\
&\quad - \exp\left(-\exp\left(-\frac{y}{\sigma}\right)\left(\exp\left(\frac{\delta_0}{\sigma}\right) + \left(\sum_{j \in \mathcal{J}} \exp\left(\frac{\delta_j}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}\right)\right).
\end{aligned}
$$

Using Result 10 yields the result. □

4.2 Empirical Example

Suppose that J firms compete in a differentiated-product market and each firm offers multiple products. The set of products that firm j produces is denoted by $\mathcal{K}_j$. Each consumer chooses an alternative from the common choice set $\mathcal{K} \equiv \bigcup_{j\in\mathcal{J}} \mathcal{K}_j$. The indirect utility of choosing alternative k_j is given by $S_{k_j} = \alpha_{k_j} + \sigma\epsilon_{k_j}$ for $k_j \in \mathcal{K}_j$, $j \in \mathcal{J}$, where ϵ_{k_j} is i.i.d. extreme value type 1. We may be interested in the expected value of indirect utility of the best alternative and the expected value of indirect utility of the best *rival* product. Let $M_{-j} = \max\{S_k,\ k \in \mathcal{K} \setminus \mathcal{K}_j\}$. Then,

Result 11′: *The conditional expectation of the best rival surplus, given the highest surplus, in model G.*

$$\mathbb{E}\left[M_{-j}|S_{k_j} = \max\{S_k, k \in \mathcal{K} \cup \{0\}\}\right] = \delta_{k_j} + \sigma \ln D_{k_j} + \sigma\gamma + \frac{1}{P_j}\left[I_{-j}(\lambda) - I(\lambda)\right],$$

where

$$I_{-j}(\lambda) = \sigma \ln \left[\exp\left(\frac{\delta_0}{\sigma}\right) + \left(\sum_{k\in\mathcal{K}\setminus\mathcal{K}_j} \exp\left(\frac{\delta_k}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}\right]$$

and P_j *is the NMNL choice probability that a consumer chooses one of the products produced by firm j:*

$$P_j = \sum\nolimits_{k_j\in\mathcal{K}_j} P_{k_j}.$$

5 GEV Sampler

This section describes a practical method to draw a random sample $\epsilon = \{\epsilon_j\}_{j=0}^{J}$ from model G. The building block for the method is an MCMC procedure developed by McFadden (1999).

Theorem (McFadden, 1999): Construct vectors ϵ^t recursively for $t = 1, ..., T$. At step t, draw (0,1) uniformly distributed random variables η^t and ζ_j^t for $j = 0, 1, ..., J$. Define $\widetilde{\epsilon}_j^t = -\sigma \log(-\log(\zeta_j^t))$, $g_t =_{j=0}^{J} (-\zeta_j^t \log(\zeta_j^t))$, and $f_t = \partial^J F(\widetilde{\epsilon}^t)/\partial\widetilde{\epsilon}_0^t ... \partial\widetilde{\epsilon}_J^t$.

Then, $\widetilde{\epsilon}^t$ is a draw from $g(\cdot)$, $g_t = g(\widetilde{\epsilon}^t)$, and $f_t = f(\widetilde{\epsilon}^t)$. Define a Markov chain

$$\epsilon^t = \begin{cases} \widetilde{\epsilon}^t & \text{if } \eta^t \leq \frac{f_t g_{t-1}}{f_{t-1} g_t} \\ \epsilon^{t-1} & \text{otherwise} \end{cases}.$$

Call the vectors ϵ^t obtained in this manner a GEV sampler. Let $f^{(t)}(\epsilon)$ denote the density of the vector ϵ^t obtained at step t of the sampler, conditioned on ϵ^0. Then, the sequence ϵ^t is an irreducible, aperiodic, Harris recurrent Markov chain, with the following properties:

1. For any ϵ^0, $\int_{-\infty}^{\infty} |f(\epsilon) - f^{(t)}(\epsilon)| d\epsilon \to 0$ as $t \to +\infty$;
2. For any real-valued function $h(\cdot)$ that is integrable with respect to f, the process is strongly ergodic, with $\frac{1}{T}\sum_{t=1}^{T} h(\epsilon^t) \to Eh \equiv \int_{-\infty}^{\infty} h(\epsilon) f(\epsilon) d\epsilon$ almost surely as $T \to +\infty$.

Thus, we can obtain a sampler, whose density is arbitrarily close to the true density. The next subsection describes how to compute the density $f_t = \partial^J F(\widetilde{\epsilon}^t)/\partial\widetilde{\epsilon}_0^t ... \partial\widetilde{\epsilon}_J^t$.

5.1 GEV density for the Nested Logit Model

Consider a generalized extreme value based nested model, with $\tilde{S}_{ij} = \delta_{ij} + \sigma\nu_{ij}$, and the CDF of $\nu_i = [\nu_{ij}]_{j\in\{0,\mathcal{J}_i\}}$ is given by

$$\begin{aligned} F(\nu_i) &= \exp\left(-\exp\left(-\frac{\nu_{i0}}{\sigma}\right) - \left(\sum_{j\in\mathcal{J}_i} \exp\left(-\frac{\nu_{ij}}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}\right) \\ &= \exp\left(-\exp\left(-\frac{\nu_{i0}}{\sigma}\right)\right) \exp\left(-\left(\sum_{j\in\mathcal{J}_i} \exp\left(-\frac{\nu_{ij}}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}\right) \\ &= A(\nu_{i0}) \times B(\nu_{ij}; j \in \mathcal{J}_i), \end{aligned}$$

where

$$\begin{aligned} A(\nu_{i0}) &= \exp\left(-\exp\left(-\frac{\nu_{i0}}{\sigma}\right)\right) \\ B(\nu_{ij}; j \in \mathcal{J}_i) &= \exp\left(-\left(\sum_{j\in\mathcal{J}_i} \exp\left(-\frac{\nu_{ij}}{\sigma(1-\lambda)}\right)\right)^{1-\lambda}\right). \end{aligned}$$

Notice - as a check - that the term B also factorizes into the product of EV(0,σ) CDFs when $\lambda = 0$.

Then,

$$\frac{\partial}{\partial \nu_{i0}} F(\nu_i) = B \frac{\partial}{\partial \nu_{i0}} A(\nu_{i0}) = f_0(\nu_{i0}) B,$$

where $f_0(\nu_{i0}) = \frac{1}{\sigma} \exp\left(-\frac{\nu_{i0}}{\sigma} - \exp\left(-\frac{\nu_{i0}}{\sigma}\right)\right)$ is the EV(0,σ) pdf.

Denote sum in the exponent of the exponential in B by C, i.e.

$$\begin{aligned} C &= C(\nu_{ij}; j \in \mathcal{J}_i) \\ &= \sum_{j \in \mathcal{J}_i} \exp\left(-\frac{\nu_{ij}}{\sigma(1-\lambda)}\right), \end{aligned}$$

so that

$$F(\nu_i) = A(\nu_{i0}) \exp\left(-C(\nu_{ij}; j \in \mathcal{J}_i)^{1-\lambda}\right).$$

Tedious calculations yield the following result: For $J_i = \#\mathcal{J}_i$,

$$\begin{aligned} f(\nu_i) &= \frac{\partial^{J_i+1}}{\prod_{j=0}^{J_i} \partial \nu_{ij}} F(\nu_i) \\ &= f_0(\nu_{i0}) \left[\prod_{j=1}^{J_i} g_0(\nu_{ij})\right] B(\nu_{ij}; j \in \mathcal{J}_i) \left[C(\nu_{ij}; j \in \mathcal{J}_i)^{-J_i\lambda} + o(\lambda) + o(\lambda^2) + \cdots + o(\lambda^{J_i-1})\right], \end{aligned} \tag{5-1}$$

where $g_0(\nu_{ij}) = \frac{1}{\sigma} \exp\left(-\frac{\nu_{ij}}{\sigma(1-\lambda)}\right)$ and $o(\lambda), o(\lambda^2), \cdots$ is a function of λ, λ^2, and so on. Notice - again as a check - that $\lambda = 0$ implies that $f(\nu_i) = \prod_{j=0}^{J_i} f_0(\nu_{ij})$.

5.2 Example

Consider the case of $J_1 = 4$, and let $B = B(\nu_{ij}; j \in \mathcal{J}_i)$ and $C = C(\nu_{ij}; j \in \mathcal{J}_i)$.

$$
\begin{aligned}
\frac{\partial}{\partial \nu_{i1}} B &= B \frac{1}{\sigma} \exp\left(-\frac{\nu_{i1}}{\sigma(1-\lambda)}\right) C^{-\lambda} \\
&= g_0(\nu_{i1}) B C^{-\lambda} \\
\frac{\partial^2}{\partial \nu_{i1} \partial \nu_{i2}} B &= g_0(\nu_{i1}) g_0(\nu_{i2}) B \left[C^{-2\lambda} + \frac{\lambda}{1-\lambda} C^{-\lambda-1} \right] \\
\frac{\partial^3}{\partial \nu_{i1} \partial \nu_{i2} \partial \nu_{i3}} B &= g_0(\nu_{i1}) g_0(\nu_{i2}) g_0(\nu_{i3}) B \left[C^{-3\lambda} + \frac{3\lambda}{1-\lambda} C^{-2\lambda-1} + \frac{\lambda(1+\lambda)}{(1-\lambda)^2} C^{-\lambda-2} \right] \\
\frac{\partial^4}{\partial \nu_{i1} \cdots \partial \nu_{i4}} B &= \left[\prod_{j=1}^{4} g_0(\nu_{ij}) \right] B \left[C^{-4\lambda} + \frac{6\lambda}{1-\lambda} C^{-3\lambda-1} + \frac{3\lambda(2\lambda+1)+\lambda(1+\lambda)}{(1-\lambda)^2} C^{-2\lambda-2} \cdots \right. \\
&\quad \left. + \frac{\lambda(1+\lambda)(2+\lambda)}{(1-\lambda)^3} C^{-\lambda-3} \right].
\end{aligned}
$$

Hence, in the case $J = 3$,

$$
f(\nu_i) = f_0(\nu_{i0}) \left[\prod_{j=1}^{3} g_0(\nu_{ij}) \right] B \left[C^{-3\lambda} + \frac{3\lambda}{1-\lambda} C^{-2\lambda-1} + \frac{\lambda(1+\lambda)}{(1-\lambda)^2} C^{-\lambda-2} \right],
$$

while

$$
\tilde{f}(\nu_i) = f_0(\nu_{i0}) \left[\prod_{j=1}^{3} g_0(\nu_{ij}) \right] B(\nu_{ij}; j \in \mathcal{J}_i) C(\nu_{ij}; j \in \mathcal{J}_i)^{-3\lambda};
$$

and in the case $J = 4$,

$$
\begin{aligned}
f(\nu_i) &= f_0(\nu_{i0}) \left[\prod_{j=1}^{3} g_0(\nu_{ij}) \right] B \left[C^{-4\lambda} + \frac{6\lambda}{1-\lambda} C^{-3\lambda-1} + \frac{3\lambda(2\lambda+1)+\lambda(1+\lambda)}{(1-\lambda)^2} C^{-2\lambda-2} \cdots \right. \\
&\quad \left. + \frac{\lambda(1+\lambda)(2+\lambda)}{(1-\lambda)^3} C^{-\lambda-3} \right],
\end{aligned}
$$

while

$$
\tilde{f}(\nu_i) = f_0(\nu_{i0}) \left[\prod_{j=1}^{4} g_0(\nu_{ij}) \right] B(\nu_{ij}; j \in \mathcal{J}_i) C(\nu_{ij}; j \in \mathcal{J}_i)^{-4\lambda};
$$

5.3 Recursion

This subsection provides a recursive characterization of $o(\lambda), o(\lambda^2), \cdots, o(\lambda^{J_i-1})$ in equation (5-1), which can be used in practice. For notational convenience, drop the index i. Let $J \geq 2$, and in the expression for $f(\nu)$ above consider the term in square brackets at level J:

$$D(J) = \gamma_{1:J} C^{-J\lambda} + \gamma_{2:J} C^{-(J-1)\lambda-1} + \cdots + \gamma_{J:J} C^{-\lambda-(J-1)},$$

where $\gamma_{1:J}, \cdots, \gamma_{J:J}$ denote the respective coefficients on the powers of C at level J which are functions of powers of λ. With this notation,

$$D(J+1) = \gamma_{1:J+1} C^{-(J+1)\lambda} + \gamma_{2:J+1} C^{-J\lambda-1} + \cdots + \gamma_{J+1:J+1} C^{-\lambda-J}.$$

Note, from the example, that

$$\begin{aligned} \gamma_{1:J} &= \gamma_{1:J+1} = 1 \\ \gamma_{J:J} &= \frac{\lambda(1+\lambda)\cdots(J-2+\lambda)}{(1-\lambda)^{J-1}} \\ \gamma_{J+1:J+1} &= \frac{\lambda(1+\lambda)\cdots(J-1+\lambda)}{(1-\lambda)^{J}}. \end{aligned}$$

The remaining coefficients at level $J+1$ can be constructed from those at level J as follows:

$$\begin{aligned} \gamma_{2:J+1} &= \frac{\gamma_{1:J} J\lambda}{1-\lambda} + \gamma_{2:J} \\ \gamma_{3:J+1} &= \frac{\gamma_{2:J}\left((J-1)\lambda+1\right)}{1-\lambda} + \gamma_{3:J} \\ &\vdots \\ \gamma_{J:J+1} &= \frac{\gamma_{J-1:J}\left(2\lambda+(J-2)\right)}{1-\lambda} + \gamma_{J:J}. \end{aligned}$$

The proof follows by induction. Consider the derivative at level J:

$$\frac{\partial^J}{\prod_{j=1}^J \partial\nu_j} B = \prod_{j=1}^J g_0(\nu_j) B\left[C^{-J\lambda} + \gamma_{2:J} C^{-(J-1)\lambda-1} + \cdots + \gamma_{J:J} C^{-\lambda-(J-1)}\right].$$

Let $D(J) = C^{-J\lambda} + \gamma_{2:J} C^{-(J-1)\lambda-1} + \cdots + \gamma_{J:J} C^{-\lambda-(J-1)}$.

Then,

$$\begin{aligned}
\frac{\partial^{J+1}}{\prod_{j=1}^{J+1}\partial\nu_j}B &= \prod_{j=1}^{J} g_0(\nu_j)\left[\left[\frac{\partial}{\partial\nu_{J+1}}B\right]D(J) + B\frac{\partial}{\partial\nu_{J+1}}D(J)\right] \\
&= \prod_{j=1}^{J+1} g_0(\nu_j)\left[BC^{-\lambda}D(J) + B\left[\frac{J\lambda}{1-\lambda}C^{-J\lambda-1} + \gamma_{2:J}\frac{(J-1)\lambda+1}{1-\lambda}C^{-(J-1)\lambda-2} + \cdots\right.\right. \\
&\quad \left.\left. +\gamma_{J:J}\frac{\lambda+J-1}{1-\lambda}C^{-\lambda-J}\right]\right] \\
&= \prod_{j=1}^{J+1} g_0(\nu_j)B\left[C^{-(J+1)\lambda} + \gamma_{2:J}C^{-J\lambda-1} + \cdots + \gamma_{J:J}C^{-2\lambda-(J-1)} + \cdots\right. \\
&\quad \left. \frac{J\lambda}{1-\lambda}C^{-J\lambda-1} + \gamma_{2:J}\frac{(J-1)\lambda+1}{1-\lambda}C^{-(J-1)\lambda-2} + \cdots + \gamma_{J:J}\frac{\lambda+J-1}{1-\lambda}C^{-C-J}\right] \\
&= \prod_{j=1}^{J+1} g_0(\nu_j)B\left[C^{-(J+1)\lambda} + \left[\gamma_{2:J} + \frac{J\lambda}{1-l\lambda}\right]C^{-J\lambda-1} + \cdots\right. \\
&\quad \left.\left[\gamma_{3:J} + \gamma_{2:J}\frac{(J-1)\lambda+1}{1-\lambda}\right]C^{-(J-1)\lambda-2} + \cdots\gamma_{J:J}\frac{\lambda+J-1}{1-\lambda}C^{-\lambda-J}\right],
\end{aligned}$$

which establishes the recursion.

6 Conclusion

This paper presents new analytical results for conditional expectations of order statistics resulting from independent GEV distributed random variables. These results are useful in many economic models of incomplete information games that are applied in demand and competition analysis, to enable comparative statics analysis and to render their estimation more efficient. The paper also offers a recursive algorithm to derive the GEV density in high dimensional problems that are common in the differentiated product literature, thereby providing an essential building block for the MCMC algorithm to simulate NMNL models proposed by McFadden (1999). Given the overwhelming popularity of the family of GEV models in applied economic analysis, these results can be expected to be useful to a wide community of applied microeconomic researchers.

References

[1] Athey, S. and P.A. Haile (2007): “Nonparametric Approaches to Auctions”, in Heckman, J.J. and E.E. Leamer (eds.): *Handbook of Econometrics*, Vol.6, Part A, chapt. 60, Amsterdam: North Holland

[2] Ausubel, L.M., Cramton, P. and R.J. Deneckere (2002): “Bargaining with incomplete information”, in Aumann, R. and S. Hart (eds.): *Handbook of Game Theory with Economic Applications*, Vol. 3, chapt. 50, Amsterdam: North Holland

[3] Beckert, W., Smith, H. and Y. Takahashi (2015): “Competitive Price Discrimination in a Spatially Differentialed Intermediate Goods Market”, unpublished manuscript

[4] Brannman, L. and L.M. Froeb (2000): “Mergers, Cartels, Set-Asides, and Bidding Preferences in Asymmetric Oral Auctions”, *The Review of Economics and Statistics*, **82(2)**, 283-290

[5] Chatterjee, K. and W. Samuelson (1983): “Bargaining under Incomplete Information”, *Operations Research*, **31(5)**, 835-851

[6] Cramton, P. (1985): “Sequential Bargaining Mechanisms”, in Roth, A.E. (ed.): *Game-theoretic Models of Bargaining*, chapt.8, Cambridge: Cambridge University Press

[7] Froeb, L.M., Tschantz, S. and P. Crooke (1998): “Mergers Among Asymmetric Bodders: A Logit Second-Price Auction Model”, unpublished manuscript

[8] David, H.A. (1981): *Order Statistics*, New York: Wiley

[9] Fudenberg, D. and J. Tirole (1983): “Sequential Bargaining with Incomplete Information”, *The Review of Economic Studies*, **50(2)**, 221-247

[10] Maskin, E. and J. Riley (2000): “Asymmetric Auctions”, *The Review of Economic Studies*, **67(3)**, 413-438

[11] McFadden, D.L. (1974): “Conditional logit analysis of qualitative choice behavior”, in P. Zarembka (ed.): *Frontiers in Econometrics*, New York: Academic Press

[12] McFadden, D.L. (1978): “Modeling the choice of residential location”, in A. Karlqvist, Lundqvist, L., Snickars, F. and J. Weibull (eds.): *Spatial Interaction Theory and Planning Models*, Amsterdam: North-Holland

[13] McFadden, D.L. (1989): “A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration”, *Econometrica*, **57(5)**, 995-1026

[14] McFadden, D.L. (1999): “Computing Willingness-to-Pay in random Utility Models”, in Melvin, J.R., Moore, J.C. and R. Riezman (eds.): *Trade, Theory and Econometrics - Essays in Honor of John S. Chipman*, New York: Routlegde

[15] Myerson, R.B. (1981): “Optimal Auction Design”, *Mathematics of Operations Research*, **6**, 58-73

[16] Myerson, R.B. (1984): “Two-Person Bargaining Problems with Incomplete Information”, *Econometrica*, **52(2)**, 461-488

[17] Myerson, R.B. and M.A. Satterthwaite (1983): “Efficient mechanisms for bilateral trading”, *Journal of Economic Theory*, **29(2)**, 265â€“281

[18] Pakes, A. and D. Pollard (1989): “Simulation and the Asymptotics of Optimization Estimators”, *Econometrica*, **57(5)**, 1027-1057

[19] Samuelson, W.F. and J.G. Riley (1981): “Optimal Auctions”, *The American Economic Review*, **71(3)**, 381-392

[20] Vickrey, W. (1961): “Counterspeculation, Auctions, and Competitive Sealed Tenders”, *The Journal of Finance*, 16(1), 8-37