Online Appendix for "Inference for Product Competition and Separable Demand"

Adam N. Smith, Peter E. Rossi, and Greg M. Allenby

A LSP vs. DP Distributions

In this section, we compare the LSP distribution to the partition distribution induced by the Dirichlet process (DP). Since the LSP distribution is parameterized by a location partition ρ_n and a dispersion parameter $\tau > 0$, we fix $n = 100$ and set ρ_n to be a partition of $K = 5$ contiguous groups with 20 elements in each. We then consider a range of values of the scale parameter: $\tau \in \{0.1, 1, 10\}$. For the DP partitioning model, we again consider different values of the concentration parameter: $\alpha \in \{0.1, 1, 10\}$. The choice of G_0 is immaterial for the purposes of studying the induced partition distribution. To show how each model's parameters affect the resulting partition distribution, we generate $10,000$ draws from each distribution and plot the associated pairwise similarity matrices in Figure 1.

As suggested by the location-scale property of the LSP distribution, more mass is placed on the location partition ρ_n as τ tends towards zero (top left). However, as τ gets large, the probability mass gets spread more evenly across $\mathcal{P}_n$ (top right). In contrast, the distribution induced by the DP has no way of being centered around a particular partition. Under this model, the probability that any two items are grouped together is therefore constant (bottom panel). The values of α only uniformly shift these probabilities for all items (i, j).

By changing the location partition to contain only one group, we can also examine how accurately the LSP model can approximate the DP model. Figure 2 replicates the LSP pairwise similarity matrices from the top panel of Figure 1, except we now let ρ_n contain

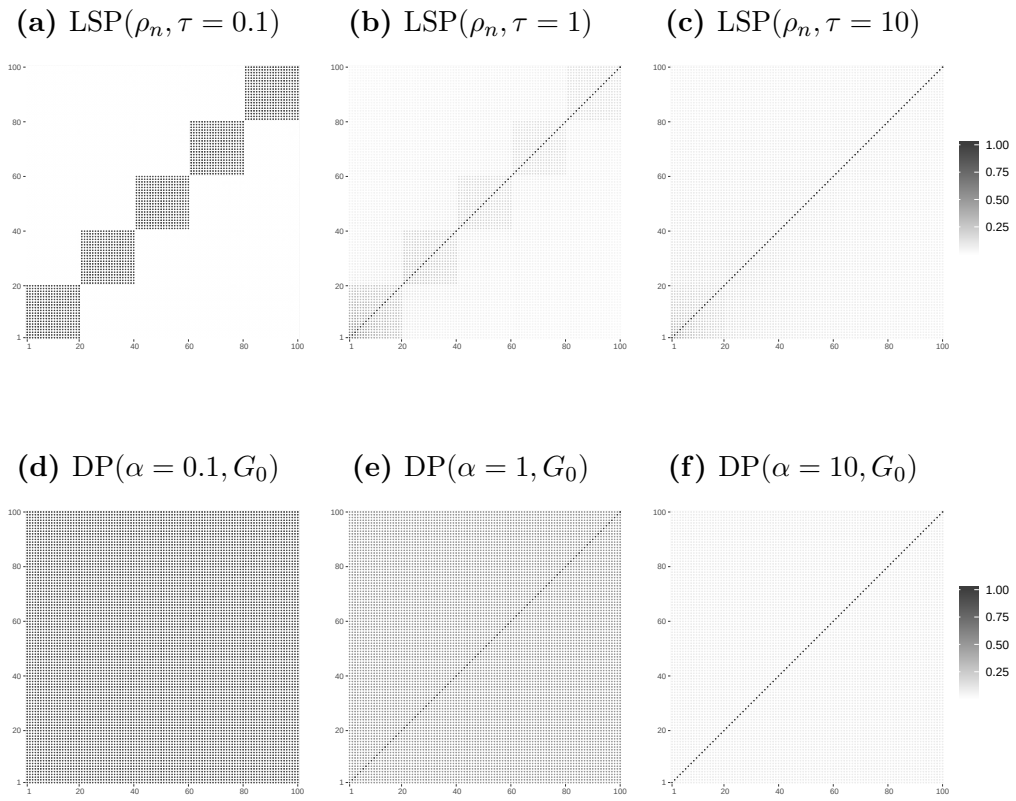


Figure 1: Pairwise similarity matrices from various LSP and DP distributions. The location partition of the LSP distribution contains five contiguous groups.

one group and divide the previous values of τ by 20. We find that the resulting pairwise similarity matrices closely match those of the DP model in the bottom panel of Figure 1.

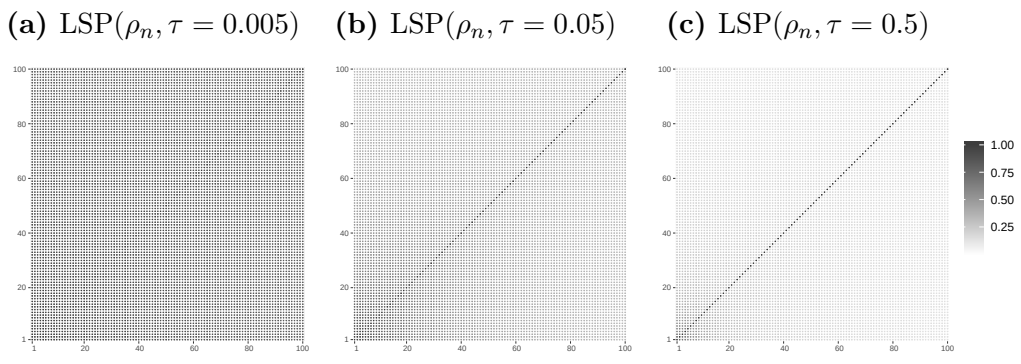


Figure 2: Pairwise similarity matrices from an LSP distribution where ρ_n contains one group and $\tau \in \{0.005, 0.05, 0.5\}$.

B Simulation Studies

B.1 Prior Sensitivity Analysis

Data are generated from the separable demand model in Section 2. Let $n = 20$ be the number of products and $T = 100$ the number of time periods. The matrix of log prices is generated from a Unif$(0, 0.5)$ distribution, the log expenditure variables are generated from a Unif$(0, 1)$ distribution, and the expenditure shares are generated from a Dirichlet$(1, \ldots, 1)$ distribution. The relatively few time periods and little variation in prices are chosen to show the impact of the prior on elasticity and partition parameters. For ease of interpretation, we let π_n be a partition of $K = 4$ contiguous groups with each group having 5 products.

$$\pi_n^{\text{true}} = (1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4)$$

Then conditional on π_n, we generate the elements of $\beta_{\pi_n} = (\eta, \theta)$: $\eta_{ii} \sim$ Unif$(-3, 0)$, $\eta_{ij}|\pi_n \sim$ Unif$(0, 3)$, and $\theta_{k\ell}|\pi_n \sim$ Unif$(-10, 10)$. We also let ψ_{ij} contain an intercept only, fix $\gamma_i = 1$ for all products, and set $\Sigma = 0.1 + I_n$. A heat map of the induced $n \times n$ price elasticity matrix is shown in Figure 3.

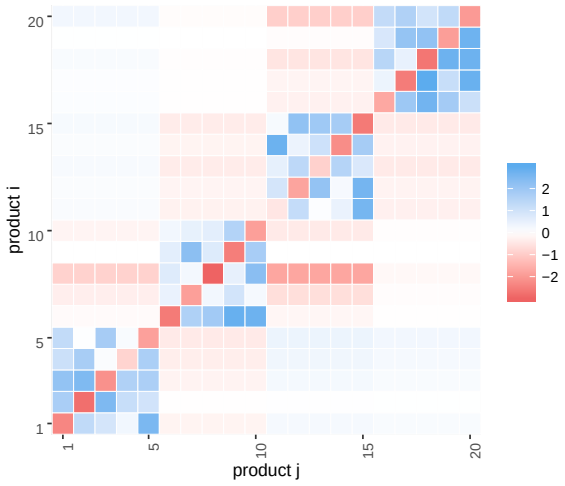


Figure 3: A heat map of the true price elasticity matrix used to generate the data.

As discussed in Section 4.3, the choice of hyperparameters in the joint prior of β_{π_n} and π_n can have significant bearing on the posterior. We highlight these effects by considering three prior specifications. The first is an informative LSP prior consistent with what is used in our empirical analysis: $\eta|\pi_n \sim \mathrm{N}(0, 10I)$, $\theta|\pi_n \sim \mathrm{N}(0, 100I)$, and $\pi_n \sim \mathrm{LSP}(\rho_n, 1/(n \log n))$ where ρ_n contains only one group. The second is more diffuse: $\eta|\pi_n \sim \mathrm{N}(0, 100I)$, $\theta|\pi_n \sim \mathrm{N}(0, 100I)$, and $\pi_n \sim \mathrm{LSP}(\rho_n, n/(n \log n))$ where again ρ_n contains only one group. The third specifies a DP prior on π_n with the concentration parameter equal to $\alpha = n$.

The model is estimated using the MCMC routine outlined in Section 4.2. We demonstrate the flexibility of our approach and treat γ as an estimated parameter with a diffuse normal prior. The Markov chain is run for $R = 50,000$ iterations and the step size of the random-walk proposals for π_n are chosen to be $v = 1/(n \log n)$.

We first consider the output from the model with an informative prior. The trace plot of the partition parameter is shown in Figure 4. The x-axis indexes the iteration, and the y-axis labels the unique partitions visited in the Markov chain. We can see that the true partition (marked by the horizontal dashed line) is reached within the first 10,000 iterations. Moreover, a small set of partitions continue to be visited throughout the course of the R iterations, while still placing highest posterior mass on the true partition.

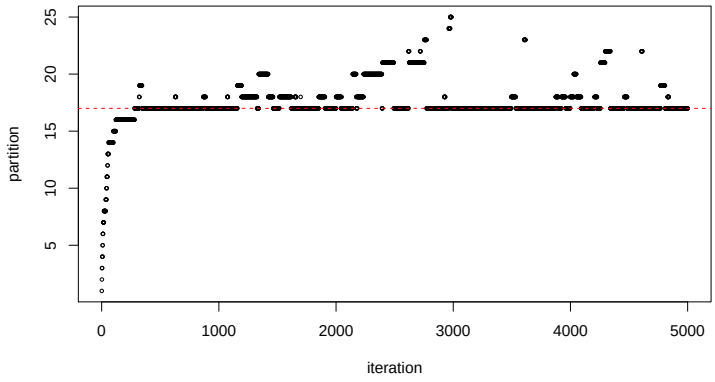


Figure 4: The trace plot for the partition parameter π_n. The dashed horizontal line represents the true partition used to generate the data.

Figure 5 plots the posterior distributions over the number of groups K corresponding to both diffuse and informative priors. We find that diffuse priors tend to favor models with many groups, as the posterior mass is concentrated concentrated to the right of $K = 4$ (shown by the dashed line). While this effect can be countered through the choice of a more informative prior on π_n, the effect would also diminish as the number of observations or amount of price variation increases. Going forward, we focus our analysis on the model with the informative LSP prior.

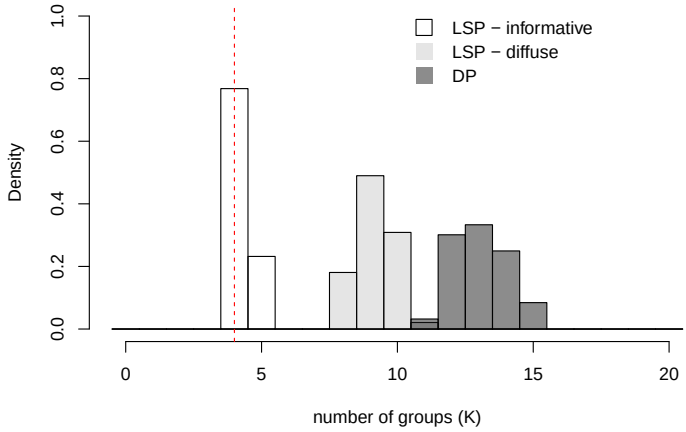
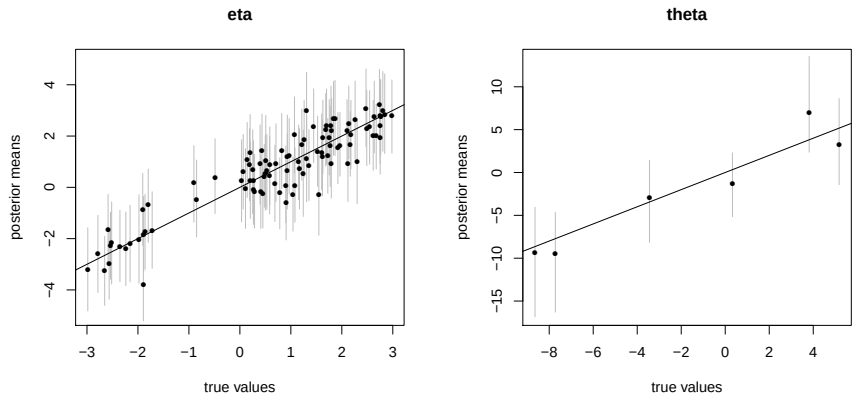


Figure 5: The posterior distributions of K are plotted for models with diffuse and informative priors. The dashed vertical line represents the true number of groups ($K = 4$).

Figure 6 plots the posterior estimates of $\beta_{\pi_n} = (\eta, \theta)$ for a model where π_n is estimated and the model where π_n is fixed at its true value. This allows us to both assess the ability of our model to recover parameters, as well as check for any losses in efficiency that arise when π_n is estimated. We find that the associated 95% credible intervals always cover the true parameter values, and there is minimal difference in the length of the credible intervals between the two models. Figure 7 plots the 95% posterior credible intervals for γ and ψ for the model when π_n is estimated. All credible intervals cover the true values. While Σ is also well recovered, we do not report the estimates for the sake of brevity.

(a) Estimated π_n



(b) Fixed π_n

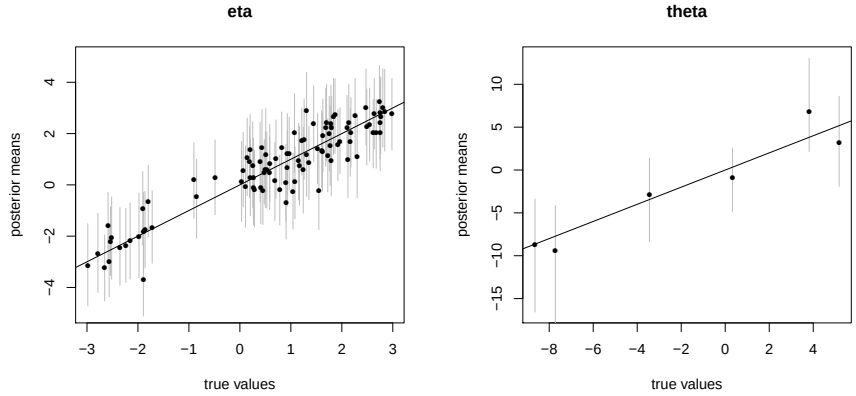


Figure 6: Conditional posterior means and 95% credible intervals are plotted against the true values of η and θ. The top and bottom panels corresponds to models where π_n is estimated and fixed, respectively. Dots indicate posterior means.

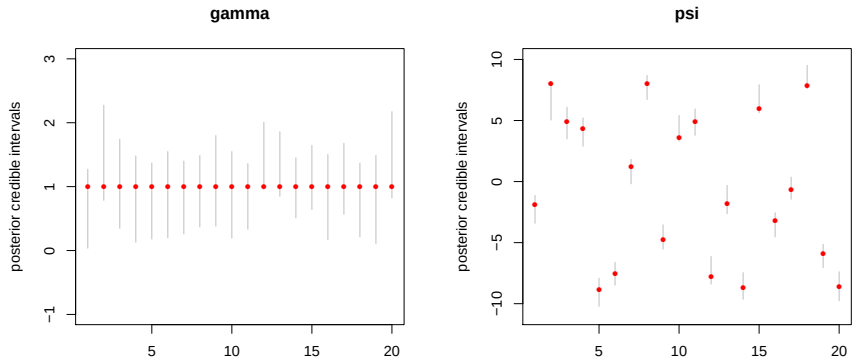


Figure 7: 95% Posterior credible intervals are plotted for γ and ψ. Dots indicate true parameter values.

B.2 High-Dimensional Example

This section examines the performance of our separable demand model with a fixed partition in high-dimensional settings. We consider data simulated from either an unrestricted log-linear model or a separable model. Let $n = 100$ denote the number of products. We explore both regular and high-dimensional settings in which T is either 75% greater than n (i.e., $T = 175$) or 75% less than n (i.e., $T = 25$), respectively. For the separable models, we let π_n be a partition with $K = 10$ contiguous groups where each group contains 10 products. Prices and expenditure are generated from a Unif(0,1) distribution and the shares are generated from a Dirichlet$(1, \ldots, 1)$ distribution. Unrestricted price elasticities are generated as $\beta_{ii} \sim$ Unif(-3, 0) and $\beta_{ij} \sim$ Unif$(0, 3)$, the separability parameters are generated as $\theta_{k\ell} \sim$ Unif$(0, 10)$, and the error covariance matrix is set to be $\Sigma = 0.1 + I_n$. Proper and relatively diffuse conjugate priors are used on all model parameters.

The results are summarized as follows: Figure 8 plots the posterior means and credible intervals for the n^2 elasticity parameters in each model; Table 1 reports in-sample and predictive fit statistics as well as the computational time associated with each model. We find that when T is large, the model that matches the data generating process performs the best. However, when T is small, the unrestricted models exhibit a tremendous amount of posterior uncertainty which render them practically useless. For example, the unrestricted model produces extremely wide 95% credible intervals that, on average, range from -35 to 35. This is also true for the unrestricted model's predictive fit, as the standard deviations of the predicted RMSE are more than 150 times larger than in the large T case. The separable model also exhibits more uncertainty when $n > T$, but to a much lesser degree. This suggests that separability can be a useful restriction for high-dimensional regression models, especially when the partitioning of goods can be reasonably specified a priori.

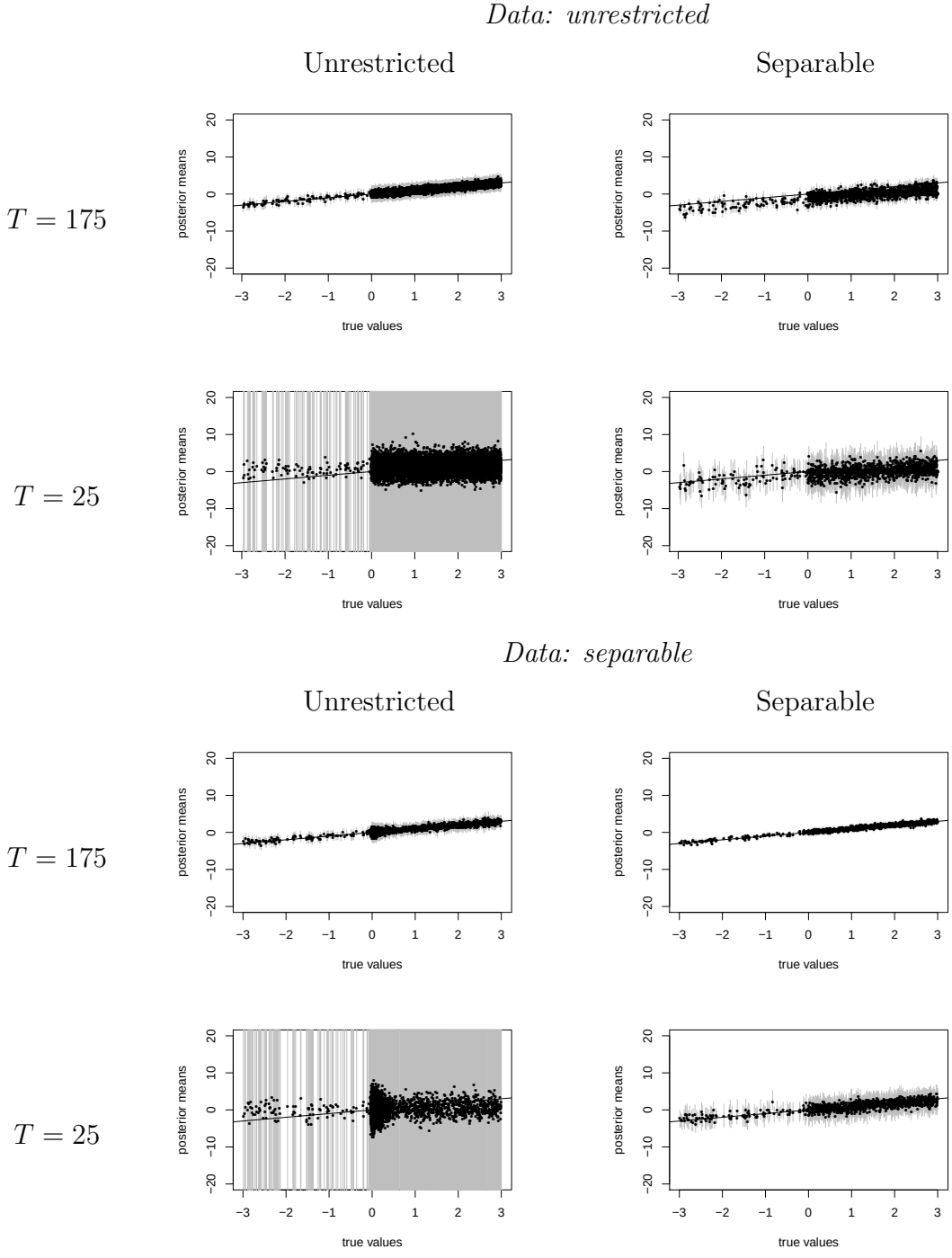


Figure 8: Posterior means and credible intervals of price elasticities for the unrestricted and separable models. The top two panels correspond to data generated from an unrestricted model, while the bottom two correspond to data generated from a separable model.

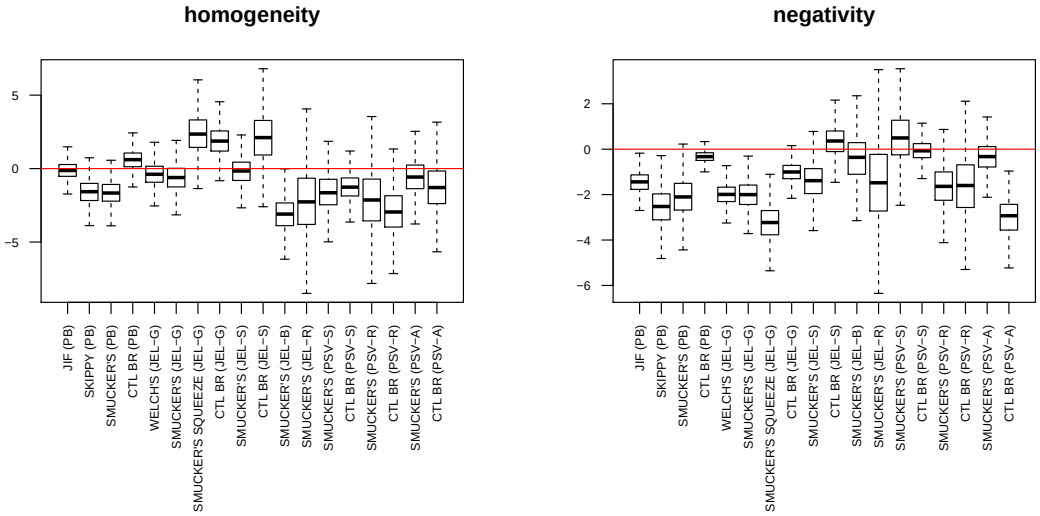
Table 1

Data	Model	In-Sample RMSE Mean	In-Sample RMSE SD	Predicted RMSE Mean	Predicted RMSE SD	Time (in sec)*
Unrestricted						
$T = 175$	Unrestricted	1.355	0.017	2.400	0.059	0.93
	Separable	71.502	1.242	70.000	1.236	30.39
$T = 25$	Unrestricted	6.123	1.294	49.949	10.76	0.75
	Separable	72.027	1.263	71.928	1.254	15.09
Separable						
$T = 175$	Unrestricted	1.358	0.015	2.433	0.053	0.90
	Separable	1.061	0.002	1.131	0.007	30.15
$T = 25$	Unrestricted	6.578	1.740	54.064	15.937	0.75
	Separable	1.318	0.032	1.830	0.043	15.15

*Time is measured as average computation time in seconds per 100 MCMC iterations.

C Integrability

(a) Jams, Jellies, and Spreads



(b) Juice Drinks

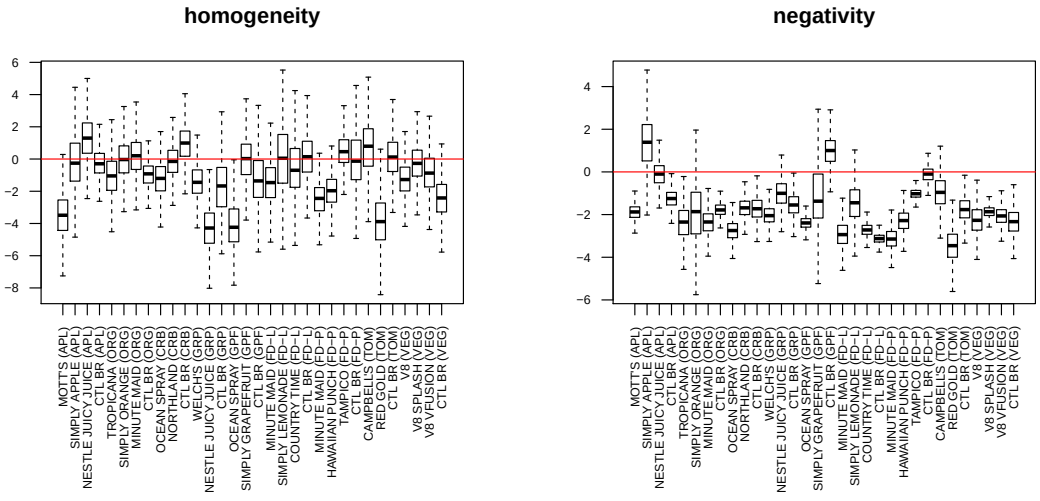


Figure 9: The distribution of the homogeneity and negativity integrability conditions are plotted for each data set. Homogeneity holds whenever the boxplot intersects the horizontal line. Negativity holds whenever the boxplot intersects or falls below the horizontal line.