

Online Appendix for:  
Triplet Embeddings for Demand Estimation

Lorenzo Magnolfi  
University of Wisconsin  
[magnolfi@wisc.edu](mailto:magnolfi@wisc.edu)

Jonathon McClure  
University of Wisconsin  
[jcmclure2@wisc.edu](mailto:jcmclure2@wisc.edu)

Alan Sorensen  
University of Wisconsin & NBER  
[sorensen@ssc.wisc.edu](mailto:sorensen@ssc.wisc.edu)

April 2024

## Appendix A IV and AIDS specifications

We discuss in this appendix two important extensions to the log-linear specification described in the paper.

### A.1 IV estimation of the log-linear model

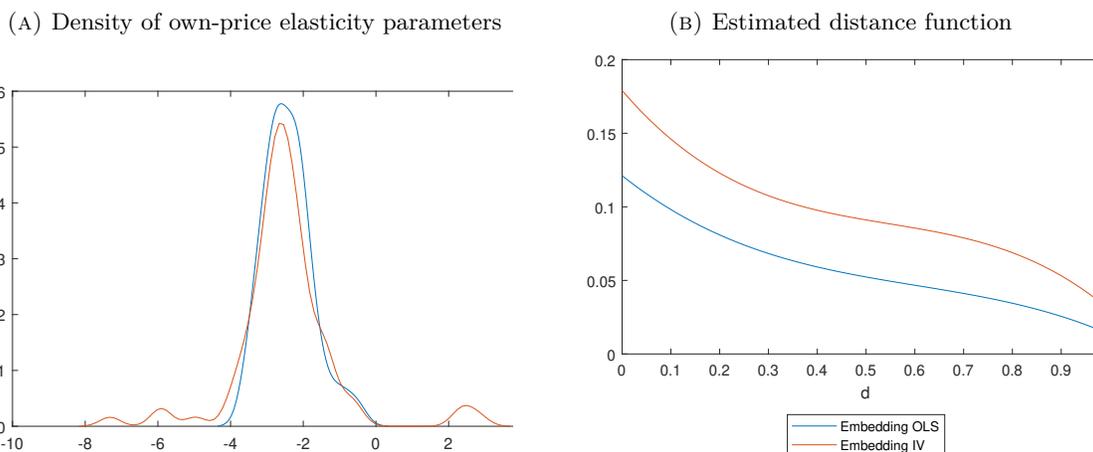
We present here results for a specification where we instrument for prices. In general, the endogeneity of prices due to the simultaneous nature of market equilibrium outcomes is a fundamental aspect of the identification of demand systems. Hence, it is helpful to discuss how to incorporate IV estimation in specifications that use embedding data. For our empirical environment, we choose to use Hausman instruments, i.e., the prices of the same goods in other markets. Similar to [Hausman and Leonard \(2007\)](#), these instruments are valid in our context of weekly data, as factors such as national advertising campaigns – which could endanger validity – are controlled for by time-fixed effects. Using instruments in unrestricted product-space demand specifications with many goods may give rise to econometric difficulties, as many instruments that vary independently are required to identify parameters ([Gandhi and Nevo, 2021](#)).

The role of distances in disciplining substitutions in our log-linear specification helps substantially: we only require instruments to identify a limited number of parameters.

To instrument for rival prices, we construct  $\sum_{k \neq j} f(d_{jk}; \gamma) \ln(z_{kt})$ , incorporating the same functional form as how rival prices are included in the regression. This adds four additional instruments for the four coefficients in  $\gamma$ .

We present in Figure 1 the results from estimating the log-linear model in Equation (1) of the paper using Hausman instruments for log price variables. The figure shows that IV estimates of own-price elasticities are comparable to the OLS estimates, although they present more outliers (including a few products with upward-sloping demand). Median price elasticity is  $-2.590$  for this specification, close to the OLS result. Despite the restrictions on the demand system made possible by the use of the embedding data, the IV estimator may still struggle to precisely identify all parameters. The distance function implied by the IV estimates has a similar shape to the one generated by the OLS estimates but is shifted upwards. Thus, IV estimates generate somewhat larger cross-elasticities in this application – the median cross-price elasticity is  $0.090$ . Overall, the demand system generated by the IV estimates is economically similar to the OLS demand system.

APPENDIX FIGURE 1: Estimates for the IV Log-linear Model



Panel A shows the density of own-price elasticities  $\beta_j$  for the log-linear model (Equation 1). Panel b shows  $f(d)$  of Equation (2) implied by the estimated  $\gamma$  parameters. OLS (IV) estimates are in blue (orange).

## A.2 AIDS specification

As another important extension, we use embedding data in a micro-founded product-space demand specification: the AIDS model of [Deaton and Muellbauer \(1980\)](#). To do so, we first transform the data to obtain products' revenue shares as  $w_j = \frac{q_j p_j}{e}$  where  $e = \sum_{k=1}^J q_k p_k$ .

The demand system is:

$$w_j = \alpha_j + \sum_{k=1 \dots J} \beta_{jk} \ln(p_k) + \theta_j \ln\left(\frac{e}{p}\right) + \epsilon_j,$$

where  $p$  is the Stone price index

$$\ln(p) = \sum_j \tilde{w}_j \ln(p_j),$$

and  $\tilde{w}_j$  is the average revenue share of product  $j$  across markets. This demand system is derived from an expenditure function that is a second-order approximation to any expenditure function (Diewert, 1971), and the demand system itself is a first-order approximation to any demand system (Deaton and Muellbauer, 1980).

Further economic properties that are normally imposed on this demand system include adding up, so that  $\sum_i \alpha_i = 0$ ,  $\sum_i \tilde{\beta}_{ij} = 0, \forall j$ ; homogeneity, or  $\sum_i \tilde{\beta}_{ji} = 0, \forall j$ ; and symmetry, or  $\beta_{ij} = \frac{1}{2}(\tilde{\beta}_{ij} + \tilde{\beta}_{ji}) = \beta_{ji}$ . An appealing feature of the AIDS demand system is that it allows the researcher to model consumers' choice problems hierarchically—that is, as a multi-stage budgeting problem (Gorman, 1959). Hence, the demand system described above can be interpreted as conditional, describing demand for a product in a certain category conditional on the expenditure in that category. However, that expenditure is also endogenous. To determine unconditional demand, one needs to also model the top-level demand equation. To do so in a scanner data context, Hausman and Leonard (2007) propose a specification:

$$\ln(Q_t) = \delta_{0t} + Z_t \theta + \delta_1 \ln(p_t) + \lambda \ln(E_t) + \eta_t$$

where  $Q_t$  is total category quantity in a certain market,  $p_t$  is the category price index, and  $E_t$  is total expenditure in market  $t$  across categories.

Incorporating distances from embeddings in the AIDS model enables us to restrict coefficients as  $\beta_{ij} = f(d_{ij})$ , or  $\beta_{ij} = \beta_i f(d_{ij})$ . Hence, the main equation of the demand system becomes:

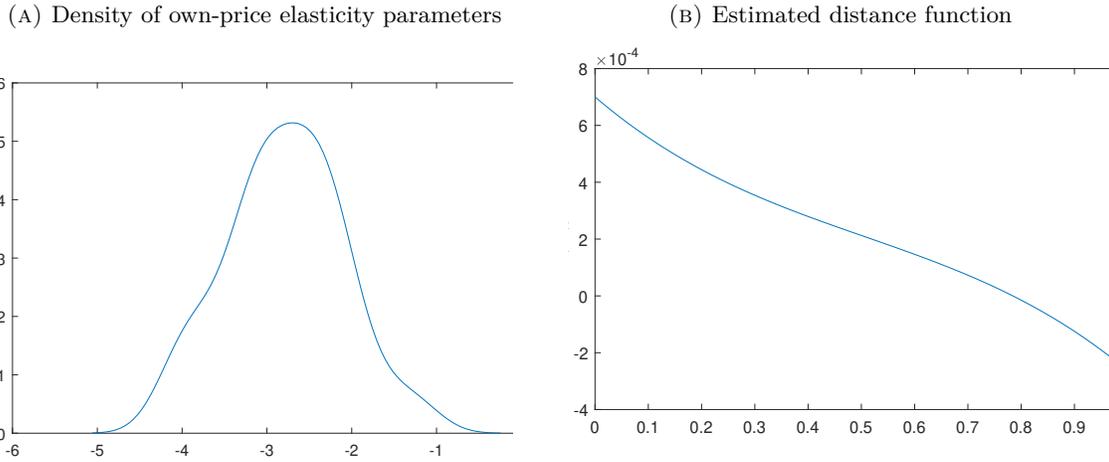
$$w_j = \alpha_j + \beta_j^{own} \ln(p_j) + \beta_j^{cross} \sum_{k=1 \dots J} f(d_{jk}) \ln(p_k) + \theta_j \ln\left(\frac{e}{p}\right) + \epsilon_j, \quad (1)$$

and the economic assumptions can be added to discipline parameters  $\beta_j$ .

We estimate Equation (1) using the data of our empirical application. Figure 2 reports estimates for this model. Overall, own-price elasticities are comparable to the log-linear

specification, with a median value of  $-2.781$ , but with a larger variance across products. While the distance function is not comparable to the one estimated for the log-linear model due to different scales, it still comes out monotonically decreasing.

APPENDIX FIGURE 2: Estimates for the AIDS Model



Panel A shows the density of own-price elasticities  $\beta_j^{own}$  for the AIDS model (Equation 4). Panel b shows  $f(d)$  of Equation (2) implied by the estimated  $\gamma$  parameters.

## Appendix B Choosing the dimensionality of the embedding

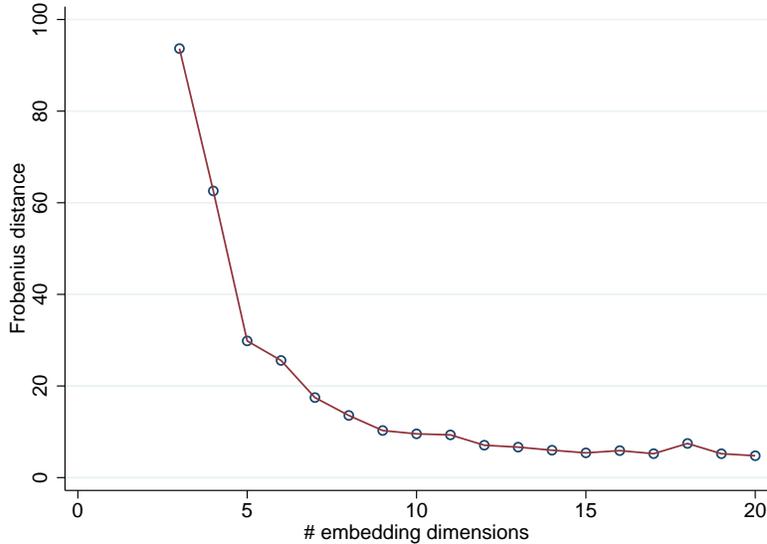
To our knowledge, there is no single, widely accepted method for choosing the dimensionality ( $m$ ) of the embedding. Embeddings are used in a wide variety of contexts, and they are estimated using many different types of data, so we expect the “right” choice of  $m$  will be application-specific. In many cases (e.g. natural language processing, image recognition) it is typical to compute embeddings with hundreds or even thousands of dimensions. In our application, we wanted to allow for enough dimensions to adequately approximate all relevant aspects of the product space, but we also wanted to keep the number of dimensions low to limit the number of random coefficients we would need to estimate in the BLP models described in Section III.B of the paper.

Some basic rules of thumb have been suggested, such as the fourth root of the number of unique categories (which in our case would yield  $m = 3$ ) or 1.6 times the square root of the number of unique categories (which in our case would yield  $m = 15$ ). Since we are using the embedding for something specific (which is to get pairwise product distances), a more hands-on approach is to compute embeddings for different values of  $m$  and choose the value of  $m$  at which the pairwise product distances settle down—which can be gauged by looking at the Frobenius distance between the matrixes of pairwise distances, or simply by calculating the correlation between the pairwise distances for successive values of  $m$ . The figure below plots the Frobenius distances for different values of  $m$ . While there is no obvious breakpoint, the curve visibly flattens beyond  $m = 8$ . If instead we rely on simple correlations, we found that the distances from a 7-dimensional embedding and a 6-dimensional embedding have a correlation of 0.99, so we concluded that increasing  $m$  beyond 6 would have little impact on our results. Later tests confirmed this: at least for the log-linear models (which are easy to re-estimate using embeddings of different dimensions), we found that our elasticity estimates hardly changed if we used higher-dimensional embeddings.

An alternative method we considered for choosing  $m$  is based on principal components analysis (PCA). Although computing an embedding is a form of dimensionality reduction, the columns of the embedding are not orthogonal. This means that PCA can be applied to the embedding itself. As a method for determining  $m$ , we considered the following algorithm:

1. Start with  $m = 1$ .
2. Iteration:
  - (a) Compute the embedding with  $m + 1$  dimensions
  - (b) Compute the principal components of the  $m + 1$  columns of the embedding

APPENDIX FIGURE 3: Convergence of product distances for higher-dimensional embeddings



The figure plots the Frobenius distance between the matrix of pairwise distances from the  $m$ -dimensional embedding and the  $(m-1)$ -dimensional embedding.

- (c) If a subset of  $k < m + 1$  principal components explain more than  $T$  percent of the variance, stop and choose the dimensionality of the embedding to be  $m$ . Otherwise, increase  $m$  and iterate.

An alternative algorithm could be performed in reverse: starting with a large  $m$ , incrementally decrease it until it takes all  $m$  principal components to explain at least  $T$  percent of the variance. The threshold  $T$  is obviously a tuning parameter that must be chosen by the researcher. In our case, if we choose a threshold of 95 percent, the algorithm above and its reverse both lead to a choice of  $m = 6$ .

## Appendix C Details about the survey and triplets sample

As noted by [Wilber et al. \(2014\)](#), the design of the comparison page trades off complexity (from the respondent’s perspective) and efficiency (from the researcher’s perspective). If the comparison page shows  $N$  products and asks the respondent to identify which  $K$  are closest to the reference, then each comparison page yields  $K(N - K)$  triplets. By choosing  $N$  to be large and  $K = N/2$ , the researcher would maximize the amount of information she gets from each comparison page. However, large values of  $N$  or  $K$  make the comparison cognitively burdensome for the respondent. The right values of  $N$  and  $K$  will be context-

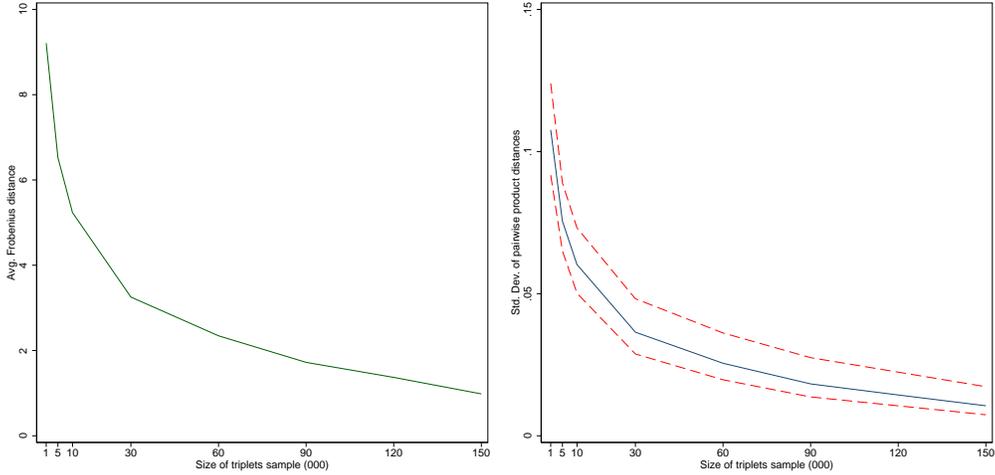
dependent; in our application, we chose  $N = 8$  and  $K = 2$  to keep the comparisons simple so that each respondent could complete many comparisons in a short amount of time.

While there are some fixed costs of designing and testing an online survey like the one we used, once it has been prepared it is surprisingly fast and cheap to collect a large sample of triplets. In our case, we got tens of thousands of triplets from undergraduate students in exchange for a few extra credit points in one of the authors' courses. When this is not an option, or in applications where we would not expect undergraduates to be well-informed about the product space, running the survey on Mechanical Turk is a natural alternative. We obtained some of our data this way, paying Turk workers \$1.10 to complete a sequence of 20 product comparison pages, which took them an average of 5 minutes. After posting the survey on Mechanical Turk, we obtained 54,348 triplets for roughly \$300 in less than 24 hours.

Another useful aspect of this method is that researchers don't need to pre-commit to a particular sample size. Additional surveys can be collected until increases in the triplets sample have little to no effect on the embedding. In our case, it appears the embedding settles down once the size of the triplet sample reaches around 100,000. From our full sample of triplets, we took 100 random subsamples of size  $N$  and recomputed the 100 corresponding embeddings. To measure the variability of the embeddings, we computed the Frobenius distances between the matrixes of pairwise product distances and the average matrix of pairwise product distances. Figure 4 shows how the average Frobenius distance (left panel) the standard deviations of the individual pairwise product distances (right panel) decline as we increase the size of the subsamples. Our final sample included over 175,000 triplets, and it appears that sampling variability in the embedding would not have been much of a concern for samples above 100,000.

As another way of evaluating sampling error in our pairwise product distances, we also computed embeddings (and associated matrixes of pairwise product distances) for 100 bootstrap resamples of our triplets data. For each of these samples, we calculated the Spearman rank correlation between the pairwise distances  $d_{jks}$  (the distance between products  $j$  and  $k$  for sample  $s$ ) and the average (across the 100 resamples) of the pairwise distances  $\bar{d}_{jk}$ . The sample that was *least* correlated with the average still had a rank correlation of 0.982. In other words, the 100 bootstrap samples all deliver almost the same ranking of pairwise product distances. We took this to imply that sampling variability from the embedding is not quantitatively important in our application. But in cases where the embedding exhibits more sensitivity to changes in the triplets sample, this bootstrap method could be used to adjust the standard errors of the estimated demand parameters.

APPENDIX FIGURE 4: Variability of product distances for subsamples of the triplets data



We took 100 random subsamples of size  $N$  (horizontal axis) and computed the matrix of pairwise product distances  $D_i$  from the resulting embeddings. The left panel plots the average of  $\|D_i - \bar{D}\|_F$ , where  $\bar{D}$  is the average of the 100 distance matrixes and  $\|\cdot\|_F$  is the Frobenius norm. The right panel plots the 10<sup>th</sup>, 50<sup>th</sup>, and 90<sup>th</sup> percentiles (across products) of the standard deviations (across subsamples) of the pairwise product distances.

## Appendix D Substitution between cereal groupings

To further explore how the use of embeddings affects estimated substitution patterns, we manually assigned the cereals in our sample to intuitive clusters and measured within- vs. between-cluster substitution. We divided cereals into the following groups: Chocolate-flavored high-sugar, Fruit-flavored high-sugar, Low-sugar, Honey-flavored, and Other. We did not base the groupings on the embedding, but they nevertheless largely correspond to different regions of the product space as measured by embedding data, and are also obviously correlated with observed characteristics. Intuitively, we expect that consumers will substitute more readily between products in the same group. We report in Table 1 below average diversion by group for each of our four models.

The results from this exercise echo previous observations. While the overall level of diversion depends on the functional form adopted, models that use embedding data predict higher average substitution to products within the same group. In contrast, models based on observed characteristics data struggle to predict more substitution to similar products in the same group.

APPENDIX TABLE 1: Mean Diversion Ratios Between Product Categories

	Observed Characteristics (Log-log)				
	Choc.	Fruit	Unsw.	Honey	Other
Chocolate	<b>0.023</b>	0.030	0.028	0.028	0.031
Fruit	0.021	<b>0.029</b>	0.023	0.025	0.027
Unsweetened	0.040	0.050	<b>0.065</b>	0.051	0.061
Honey	0.036	0.045	0.052	<b>0.046</b>	0.049
Other	0.032	0.041	0.044	0.039	<b>0.044</b>
	Embeddings (Log-log)				
	Choc.	Fruit	Unsw.	Honey	Other
Chocolate	<b>0.033</b>	0.030	0.029	0.029	0.032
Fruit	0.022	<b>0.040</b>	0.023	0.028	0.029
Unsweetened	0.045	0.053	<b>0.070</b>	0.063	0.071
Honey	0.038	0.045	0.054	<b>0.055</b>	0.056
Other	0.031	0.038	0.044	0.042	<b>0.049</b>
	Observed characteristics (BLP)				
	Choc.	Fruit	Unsw.	Honey	Other
Chocolate	<b>0.004</b>	0.006	0.006	0.006	0.006
Fruit	0.004	<b>0.006</b>	0.006	0.006	0.006
Unsweetened	0.004	0.006	<b>0.006</b>	0.006	0.006
Honey	0.004	0.006	0.006	<b>0.006</b>	0.006
Other	0.004	0.006	0.006	0.006	<b>0.006</b>
	Embeddings (BLP)				
	Choc.	Fruit	Unsw.	Honey	Other
Chocolate	<b>0.011</b>	0.004	0.002	0.006	0.007
Fruit	0.003	<b>0.009</b>	0.005	0.008	0.005
Unsweetened	0.001	0.005	<b>0.010</b>	0.003	0.007
Honey	0.006	0.008	0.004	<b>0.008</b>	0.005
Other	0.004	0.005	0.006	0.004	<b>0.007</b>

Diversion values are mean diversion ratio to products within each category, read across rows. So, for example, in the Characteristics Log-log model, the average diversion between Fruit-flavored high-sugar cereals is 2.9 percent, whereas the average diversion from Fruit-flavored high-sugar cereals to Unsweetened cereals is 2.3 percent.

## Appendix E Additional figures and tables

APPENDIX FIGURE 5: Survey intro page

**Cereal Survey**

We are collecting data on the similarity of different kinds of breakfast cereal sold in the United States. You will see a sequence of 20 pages asking you to indicate which two cereals are most similar to a reference brand. Even if none of the cereals are very similar to the reference brand, you will need to choose the two you think are the most similar.

Before proceeding, please answer the following questions:

**What is your 10-digit student ID?**

**How often do you eat breakfast cereal?**

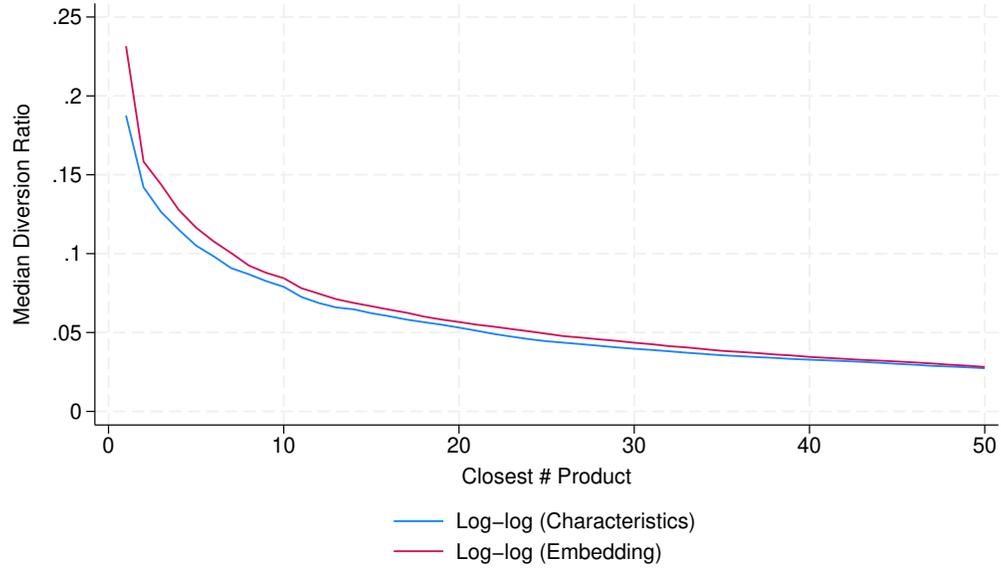
- Less than once per week
- Once or twice per week
- Three or more times per week

**How many different breakfast cereals have you personally tried while living in the United States?**

- None
- Between one and ten
- More than ten

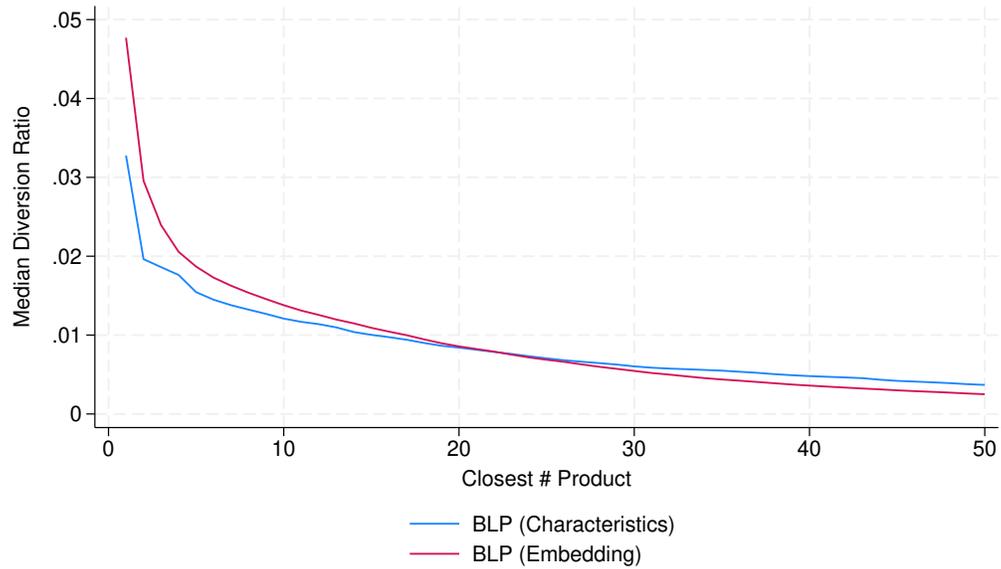
Survey respondents completed this preliminary survey before seeing the product comparison pages.

APPENDIX FIGURE 6: Diversion by closest product rank for the log-linear model



The figure shows the median (across all markets) diversion by closest product rank for the log-linear model estimated with observable characteristics (in blue) and embedding data (in red).

APPENDIX FIGURE 7: Diversion by closest product rank for the BLP model



The figure shows the median (across all markets) diversion by closest product rank for the BLP model estimated with observable characteristics (in blue) and embedding data (in red).

APPENDIX TABLE 2: Comparison of Baseline and Flexible Distance Results

Parameter	Variable	Baseline (1)		Flexible (2)	
		Estimate	SE	Estimate	SE
$\gamma$	$d_{jk}^0$	0.121	0.002	0.123	0.002
	$d_{jk}^1$	-0.263	0.013	-0.293	0.011
	$d_{jk}^2$	0.346	0.026	0.440	0.023
	$d_{jk}^3$	-0.191	0.016	-0.257	0.015
$\omega_m$	$x_{j1}$	1	—	1	—
	$x_{j2}$	1	—	0.983	0.052
	$x_{j3}$	1	—	0.000	0.019
	$x_{j4}$	1	—	1.791	0.102
	$x_{j5}$	1	—	1.592	0.083
	$x_{j6}$	1	—	0.994	0.055
Observations		684,476		684,476	
R-squared		0.838		0.973	
Median Own-Elast		-2.486		-2.475	
Median Cross-Elast		0.051		0.054	

The table reports estimates of  $\gamma$  and  $\omega_m$  parameters from Equations (2) and (3). Column 1 refers to the baseline specification of the model, which uses Euclidean distances in the  $f$  function. Column 2 refers to the flexible model using the specification of Equation (3).

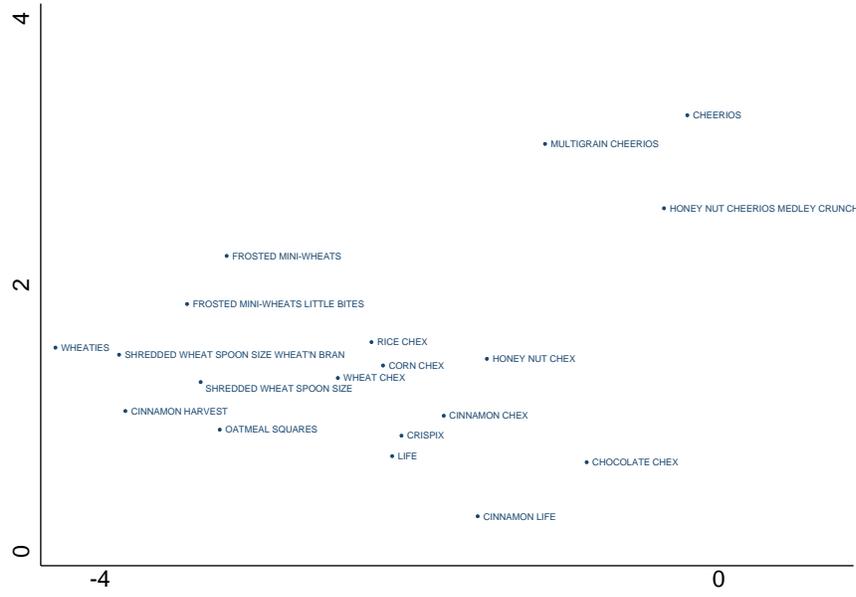
APPENDIX TABLE 3: Estimated coefficients of mixed embedding BLP model

Parameter	Variable		Parameter	Interaction	
$\beta$	Price	-1.323 (0.382)			
$\Sigma$	Constant	5.882 (0.274)	$\Pi$	Income	Kids
	Price	0.857 (0.039)		0.002 (0.036)	-0.318 (0.071)
	$x_1$	0.115 (0.072)		-0.023 (0.026)	-
	$x_2$	0.164 (0.138)		0.076 (0.029)	-
	$x_3$	0.000 (0.181)		-0.065 (0.024)	-
	$x_4$	0.000 (0.242)		0.103 (0.019)	-0.030 (0.028)
	$x_5$	0.177 (0.348)		-0.116 (0.020)	0.059 (0.036)
	$x_6$	1.3283 (0.159)		0.100 (0.026)	-0.024 (0.034)
	$x_7$	0.060 (0.239)		0.101 (0.030)	-0.080 (0.037)
	$x_8$	0.000 (0.335)		-0.032 (0.023)	0.114 (0.034)
	$x_9$	0.980 (0.088)		-0.060 (0.022)	-0.114 (0.035)
	Non-linear Variables	Mixed			
	Median Own-price Elasticity	-2.441			
	Median Outside Diversion	0.235			

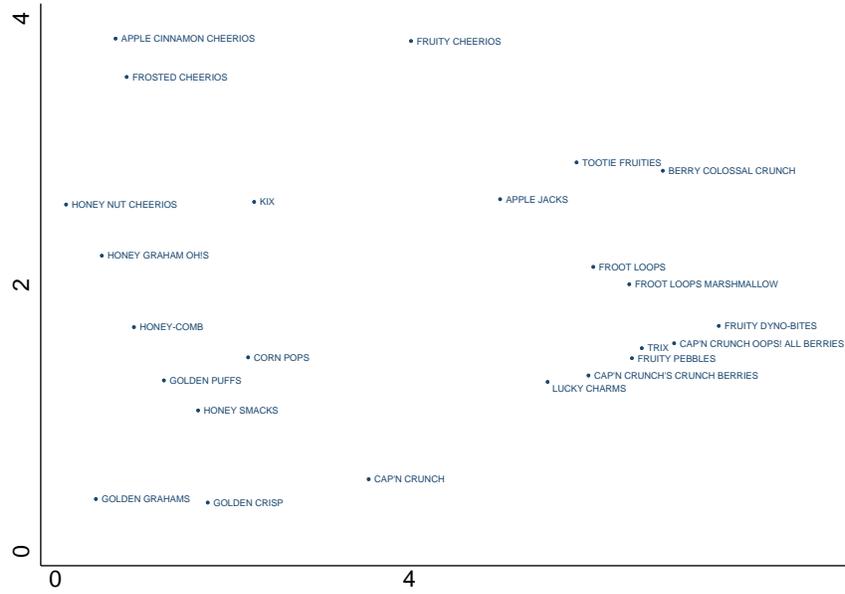
The table reports estimates (on top) and standard errors (below) for the parameters of the mixed embedding BLP model. Characteristics  $x_1$  through  $x_9$  refer to sugar, fiber, calories from fat, and a 6D embedding.  $n = 32,385$ .

Figures 8 through 11 show finer detail of the two-dimensional embedding shown in Figure 2 of the paper, breaking the plot into four regions so that all of the 86 cereal brands in our sample can be shown.

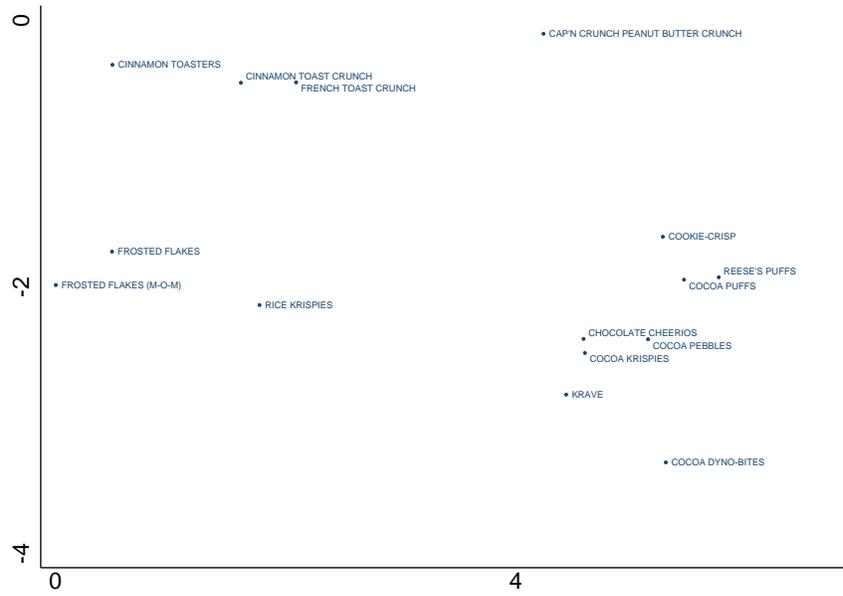
APPENDIX FIGURE 8: 2-D Embedding: Northwest quadrant



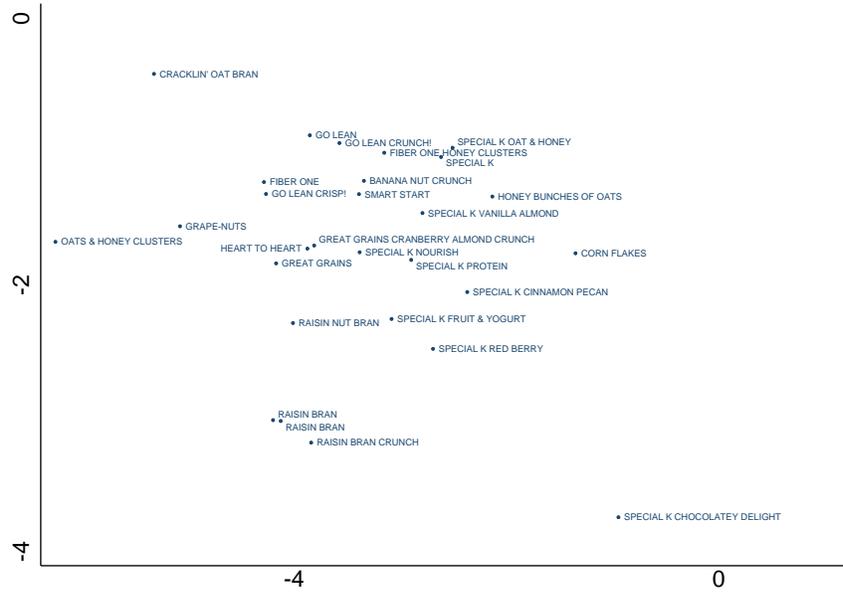
APPENDIX FIGURE 9: 2-D Embedding: Northeast quadrant



APPENDIX FIGURE 10: 2-D Embedding: Southeast quadrant



APPENDIX FIGURE 11: 2-D Embedding: Southwest quadrant



## References

- DEATON, A. AND J. MUELLBAUER (1980): “An almost ideal demand system,” *The American Economic Review*, 70, 312–326.
- DIEWERT, W. E. (1971): “An application of the Shephard duality theorem: A generalized Leontief production function,” *Journal of Political Economy*, 79, 481–507.
- GANDHI, A. AND A. NEVO (2021): “Empirical models of demand and supply in differentiated products industries,” in *Handbook of Industrial Organization*, Elsevier, vol. 4, 63–139.
- GORMAN, W. M. (1959): “Separable utility and aggregation,” *Econometrica*, 469–481.
- HAUSMAN, J. A. AND G. K. LEONARD (2007): “Estimation of patent licensing value using a flexible demand specification,” *Journal of Econometrics*, 139, 242–258.
- WILBER, M., I. KWAK, AND S. BELONGIE (2014): “Cost-effective hits for relative similarity comparisons,” in *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, vol. 2, 227–233.