Supplementary Appendix to "Sparse demand systems: corners and complements"

Arthur Lewbel

Boston College and IFS

and

Lars Nesheim

CeMMAP, IFS and UCL

September 2019

Abstract

This Supplementary Appendix presents technical details for the paper "Sparse demand systems: corners and complements." These include details of the hyperspherical transformation, the log likelihood function, and the hedonic price estimation. It also presents some summary statistics for the data.

Correspondence: l.nesheim@ucl.ac.uk

Acknowledgement: We gratefully acknowledge financial support from the ESRC through the ESRC Centre for Microdata Methods and Practice (CeMMAP) grant number ES/I034021/1 and the European Research Council (ERC) under ERC-2009-AdG grant agreement number 249529. Data supplied by Kantar Worldpanel. The use of Kantar Worldpanel data in this work does not imply the endorsement of Kantar Worldpanel in relation to the interpretation or analysis of the data. All errors and omissions remained the responsibility of the authors.

1 Introduction

This Supplementary Appendix presents technical details for the paper "Sparse demand systems: corners and complements." Appendix A describes the hyperspherical coordinate representation used in the paper. Appendix B derives the log likelihood function and presents algebraic manipulations that are used to compute the value of the log likelihood. Appendix C presents additional summary statistics for the data. Appendix D presents details of the hedonic price functions estimated.

A Hyperspherical representation of B

As discussed in Section 5.2 in the paper, it is convenient to reparameterize the matrix B in hyperspherical coordinates. This representation is derived as follows. Since B is upper triangular, $b_{kj} = 0$ if $k > j$. The number of nonzero elements in column B_j is $\overline{k} = \min\{K, j\}$. Let $C_j = \left[c_{1j}, ..., c_{\overline{k}-1}\right]^T$. The hyperspherical coordinate representation of the nonzero elements of B_j is given by $(d_j, C_j) = H(B_j)$ where H^{-1} is defined by

$$\begin{aligned}
B(1,j) &= d_j \cos(c_{1j}) \\
B(2,j) &= d_j \sin(c_{1j}) \cos(c_{2j}) \\
B(3,j) &= d_j \sin(c_{1j}) \sin(c_{2j}) \cos(c_{3j}) \\
&\vdots \\
B\left(\overline{k}-1,j\right) &= d_j \sin c_{1j} \cdots \sin\left(c_{\overline{k}-2}\right) \cos\left(c_{\overline{k}-1}\right) \\
B\left(\overline{k},j\right) &= d_j \sin(c_{1j}) \cdots \sin\left(c_{\overline{k}-2}\right) \sin\left(c_{\overline{k}-1}\right)
\end{aligned} \tag{A.1}$$

with $d_j > 0$, $c_{kj} \in [0, \pi]$ for $k < \overline{k} - 2$ and $c_{\overline{k}-1} \in [0, 2\pi)$.

B Estimation details

In this section we derive the components of the log likelihood function for 3 cases. Case 1 applies to observations in which a household purchased K goods. Case 2 applies to observations in which a household bought more than zero and fewer than K goods. Case 3 applies to observations in which a household bought zero goods.

B.1 Case 1: choice of K goods

The notation is the same as the main paper as defined in Section 3 and in Section 5.2

We drop household subscripts h to ease notation.

Suppose the goods are sorted so that $q = (q_1, 0)$. Let $p = (p_1, p_2)$ be the corresponding vector of prices. That is, the first K elements are non-negative and the remaining $J - K$ elements are 0. Let $B = [B_1\ B_2]$ as in Section 3.2.

Inverting the demand function given in equation (3.6) in Section 3.2 in the paper, inverse demand is

$$\begin{aligned} e &= \left(B_1^T\right)^{-1}\left(p_1 + B_1^T B_1 q_1\right) \\ &= \left(B_1^T\right)^{-1} p_1 + B_1 q_1 \\ p_2 &\geq B_2^T \left(B_1^T\right)^{-1} p_1. \end{aligned}$$

Since B is a function of η, $\eta \sim N(0, I)$ and $e \sim N(\mu, \Sigma)$, the case 1 log-likelihood is

$$\ln f_1(q, p, \theta) = \int\limits_{\eta} \left\{ \ln \phi \left[\left(B_1^T\right)^{-1} p_1 + B_1 q_1, \mu, \Sigma \right] + \ln\left(\det\left(B_1\right)\right) \right\} \phi(\eta, 0, I) d\eta$$

where f_1 is the case 1 density of q conditional on p and ϕ is the normal density function. Note that parameter values must satisfy the constraints that $p_2 \geq B_2^T \left(B_1^T\right)^{-1} p_1$.

B.2 Case 2: Choice of fewer than K goods

We first derive the likelihood function for fixed B.

Suppose a household chooses $q = (q_1, 0)$ with $q_1 > 0$ and $\dim(q_1) = d_1 < K$. In this case, for each q_1, there are multiple vectors e that satisfy the first order conditions

$$-p_1 - B_1^T (B_1 q_1 - e) = 0 \tag{B.1}$$

$$-p_2 - B_2^T (B_1 q_1 - e) \leq 0 \tag{B.2}$$

$$q_1 > 0. \tag{B.3}$$

In fact, the set of e values satisfying the first order conditions is a linear space of dimension $K - d_1$. In these expressions, B_1 is a $K \times d_1$ matrix with $d_1 < K$ and B_2 is a $(K \times J - d_1)$ matrix.

Let

$$B_1 = USV^T$$

be the singular value decomposition of B_1 where U is orthogonal $(K \times K)$, $S = \begin{bmatrix} S_1 \\ 0 \end{bmatrix}^T$ where S_1 is diagonal $(d_1 \times d_1)$ and V is orthogonal $(d_1 \times d_1)$. Define $\widetilde{e} = U^T e$ and partition $\widetilde{e} = (\widetilde{e}_1, \widetilde{e}_2)$ where $\widetilde{e}_1$ is $(d_1 \times 1)$ and $\widetilde{e}_2$ is $(d_2 \times 1)$. Then rewrite $(B.1)$ as

$$V \begin{bmatrix} S_1 & 0 \end{bmatrix} \begin{bmatrix} \widetilde{e}_1 \\ \widetilde{e}_2 \end{bmatrix} = p_1 + B_1^T B_1 q_1$$

or

$$V S_1 \widetilde{e}_1 = p_1 + B_1^T B_1 q_1. \tag{B.4}$$

For each q_1 there are multiple vectors $\widetilde{e}$ that solve $(B.4)$. In fact, there is a linear space of dimension d_2. In other words, for each $(q_1, \widetilde{e}_2) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, there is a unique $\widetilde{e}_1$ defined by

$$\widetilde{e}_1 = G_0 p_1 + G_1 q_1 \tag{B.5}$$

where

$$\begin{aligned} G_0 &= S_1^{-1} V^T \\ G_1 &= S_1^{-1} V^T \left(B_1^T B_1 \right). \end{aligned} \tag{B.6}$$

Since B_1 has rank d_1 by assumption, S_1 is a $(d_1 \times d_1)$ invertible diagonal matrix and by construction $V^{-1} = V^T$.

Since

$$\widetilde{e} = U^T e$$

$\widetilde{e} \sim N\left(\widetilde{\mu}, \widetilde{\Sigma}\right)$ where $\widetilde{\mu} = U^T \mu$ and $\widetilde{\Sigma} = U^T \Sigma U$. Consider the partially observed random vector $(q_1, \widetilde{e}_2)$. q_1 is observed but $\widetilde{e}_2$ is not. The expressions above imply that the density of $(q_1, \widetilde{e}_2)$ is

$$f_{q_1 \widetilde{e}_2}(q_1, \widetilde{e}_2) = f_{\widetilde{e}}(G_0 p_1 + G_1 q_1, \widetilde{e}_2) \cdot \det(G_1)$$

where (G_0, G_1) are defined in $(B.6)$.

We observe q_1 if inequality $(B.2)$ is satisfied. Since $B_1 = USV^T$ and $e = U\widetilde{e}$, this is equivalent to

$$-p_2 - B_2^T U \left(SV^T q_1 - \widetilde{e} \right) \leq 0. \tag{B.7}$$

Partitioning $\widetilde{B}_2 = U^T B_2$ $(K \times J - d_1)$ as

$$\widetilde{B}_2 = \begin{bmatrix} \widetilde{B}_{21} \\ \widetilde{B}_{22} \end{bmatrix}$$

where $\widetilde{B}_{21}$ is size $(d_1 \times J - d_1)$ and $\widetilde{B}_{22}$ is size $(d_2 \times J - d_1)$, inequality $(B.7)$ is

$$-p_2 - \begin{bmatrix} \widetilde{B}_{21}^T & \widetilde{B}_{22}^T \end{bmatrix} \left(\begin{bmatrix} S_1 V^T q_1 \\ 0 \end{bmatrix} - \begin{bmatrix} \widetilde{e}_1 \\ \widetilde{e}_2 \end{bmatrix} \right) \leq 0$$

or

$$-p_2 - \widetilde{B}_{21}^T \left(S_1 V^T q_1 - \widetilde{e}_1 \right) + \widetilde{B}_{22}^T \widetilde{e}_2 \leq 0$$

Substituting from equation $(B.5)$ this is equivalent to

$$\widetilde{B}_{22}^T \widetilde{e}_2 \leq p_2 - \widetilde{B}_{21}^T G_0 p_1 + \widetilde{B}_{21}^T \left(S_1 V^T - G_1 \right) q_1. \tag{B.8}$$

Rewrite $(B.8)$ as

$$M_1 \widetilde{e}_2 \leq M_2$$

where

$$M_1 = \widetilde{B}_{22}^T$$

is a $(J - d_1 \times d_2)$ matrix and

$$M_2 = p_2 - \widetilde{B}_{21}^T G_0 p_1 + \widetilde{B}_{21}^T \left(S_1 V^T - G_1 \right) q_1$$

is $(J - d_1 \times 1)$.

Then the Case 2 likelihood, conditional on $B(\eta)$ and p is

$$f_2 \left[q, p, B(\eta), \theta \right] = \int f_{q_1 \widetilde{e}_2} \left(q_1, \widetilde{e}_2 \right) 1 \left(M_1 \widetilde{e}_2 \leq M_2 \right) d\widetilde{e}_2. \tag{B.9}$$

Note that $f_2 \left[q, p, B(\eta), \theta \right] = 0$ if $\Pr \left(M_1 \widetilde{e}_2 \leq M_2 \right) = 0$.

Let $d_2 = K - d_1$, let $\widetilde{\Sigma}_{22} = \widetilde{C}_2^T C_2$ be the variance of $\widetilde{e}_2$. That is $\widetilde{C}_2^T$ is the upper triangular cholesky decomposition of $\widetilde{\Sigma}_{22}$. Define $\widetilde{e}_2 = \widetilde{C}_2^T z_2 + \widetilde{\mu}_2$ and note that after a change of variables the density of $\widetilde{e}$ can be written

$$f_{\widetilde{e}}\left(\widetilde{e}_1, z_2\right) = f_{\widetilde{e}_1}\left(\widetilde{e}_1, \nu_1\left(z_2\right), \Omega_1\right) \frac{e^{-0.5 z_2^T z_2}}{(2\pi)^{\frac{d_2}{2}}}$$

where $\widetilde{e}_1 \sim N\left(\nu_1, \Omega_1\right)$ and $z_2 \sim N\left(0, I\right)$ where

$$\begin{aligned} v_1 &= \widetilde{\mu}_1 + \widetilde{\Sigma}_{12} \widetilde{C}_2^{-1} z_2 \\ \Omega_1 &= \widetilde{\Sigma}_{11} - \widetilde{\Sigma}_{12} \widetilde{\Sigma}_{22}^{-1} \widetilde{\Sigma}_{21}. \end{aligned}$$

Therefore, $(B.9)$ can be written

$$f_2\left[q, p, B(\eta), \theta\right] = \int f_{q_1 z_2}\left(q_1, z_2\right) 1\left(\widetilde{M}_1 z_2 \leq \widetilde{M}_2\right) dz_2 \tag{B.10}$$

where

$$\begin{aligned} f_{q_1 z_2}\left(q_1, z_2\right) &= f_{e_1|z_2}\left(G_0 + G_1 q_1, \nu_1\left(z_2\right), \Omega_1\right) \frac{e^{-0.5 z_2^T z_2}}{(2\pi)^{\frac{d_2}{2}}} \\ &= \widetilde{f}_{q_1 z_2}\left(q_1, z_2\right) \frac{e^{-0.5 z_2^T z_2}}{(2\pi)^{\frac{d_2}{2}}} \end{aligned}$$

$$\begin{aligned} \widetilde{M}_1 &= M_1 \widetilde{C}_2^T \\ \widetilde{M}_2 &= M_2 - M_1 \widetilde{\mu}_2 \end{aligned}$$

The matrix $\widetilde{M}_1$ has the QR decomposition

$$\widetilde{M}_1 = RQ$$

where R is $(J - d_1 \times d_2)$ lower triangular and Q is $(d_2 \times d_2)$ orthogonal. Then using the change of variable $z_2 = Q^{-1}x$, the integral can be written as

$$f_2\left[q, p, B(\eta), \theta\right] = \int\limits_{RQz_2 \leq D} \widetilde{f}_{q_1 z_2}\left(q_1, z_2\right) \frac{e^{-0.5 z_2^T z_2}}{(2\pi)^{\frac{d_2}{2}}} dz_2 \quad \text{(B.11)}$$

$$= \int\limits_{Rx \leq D} \widetilde{f}_{q_1 z_2}\left(q_1, Q^{-1}x\right) \frac{e^{-0.5 x^T x}}{(2\pi)^{\frac{d_2}{2}}} dx \quad \text{(B.12)}$$

since Q is an orthogonal matrix. (That is $Q^{-1}Q = I$ and $\det(Q) = 1$) The matrix R is lower triangular. Therefore, row i has at most i nonzero elements.

Start from x_{d_2}. Let $J_{d_2}^+$ be the set of rows of R that have positive elements in column d_2 and $J_{d_2}^-$ the set with negative elements. Then for all $j \in J_{d_2}^+$,

$$-\infty \leq x_{d_2} \leq \frac{D_j - \sum\limits_{i<d_2} R(j,i)\, x_i}{R(j, d_2)}$$

and for all $j \in J_{d_2}^-$,

$$\frac{D_j - \sum\limits_{i<d_2} R(j,i)\, x_i}{R(j, d_2)} \leq x_{d_2} \leq \infty.$$

So, the bounds on x_{d_2} are $x_{d_2} \in \left[x_{d_2}^L, x_{d_2}^H\right]$ where

$$x_{d_2}^L = \max\left(-\infty, \max_{j \in J_{d_2}^-}\left(\frac{D_j - \sum\limits_{i<d_2} R(j,i)\, x_i}{R(j, d_2)}\right)\right)$$

and

$$x_{d_2}^H = \min\left(\infty, \min_{j \in J_{d_2}^+}\left(\frac{D_j - \sum\limits_{i<d_2} R(j,i)\, x_i}{R(j, d_2)}\right)\right).$$

We repeat the calculation for $j = d_2 - 1$ through 1. Then the integral is

$$f_2[q,p,B(\eta),\theta] = \int\limits_{x_1^L}^{x_1^H} \cdots \int\limits_{x_{d_2}^L}^{x_{d_2}^H} \widetilde{f}_{q_1 z_2}\left(q_1, Q^{-1}x\right) \frac{e^{-0.5x^T x}}{(2\pi)^{\frac{d_2}{2}}} dx. \tag{B.13}$$

Next for all $j \leq d_2$ define $u_j = \Phi(x_j)$. Then making the change of variables, the integral is equivalent to

$$f_2[q,p,B(\eta),\theta] = \int\limits_{u_1^L}^{u_1^H} \cdots \int\limits_{u_{d_2}^L}^{u_{d_2}^H} \widetilde{f}_{q_1 z_2}\left(q_1, Q^{-1}x(u)\right) du \tag{B.14}$$

where

$$\begin{aligned} u_j^L &= \Phi\left(x_j^L\right) \\ u_j^H &= \Phi\left(x_j^H\right). \end{aligned}$$

Finally, for all $j \leq d_2$ making the change of variable $u_j = \frac{\left(u_j^H - u_j^L\right)(1+v_j)}{2}$, this is equivalent to

$$f_2[q,p,B(\eta),\theta] = \int\limits_{-1}^{1} \cdots \int\limits_{-1}^{1} \prod_{j=1}^{d_2} \left(\frac{u_j^H - u_j^L}{2}\right) \widetilde{f}_{q_1 z_2}\left(q_1, Q^{-1}x(v)\right) dv. \tag{B.15}$$

This equals 0 if $u_j^H \leq u_j^L$ for any j.

The conditional density function f_2 depends on the parameters θ and on the random coefficient η. Integrating out the random coefficients, the Case 2 likelihood function is

$$\ln f_2(q,p,\theta) = \int\limits_{\eta} f_2[q,p,B(\eta),\theta]\,\phi(\eta) d\eta.$$

B.3 Case 3: Choice of 0 goods

Suppose a household chooses $q = 0$. In this case, the first-order conditions are

$$-p + B^T e \leq 0. \tag{B.16}$$

In this inequality, B is a $K \times J$ matrix. Rewrite the inequality as

$$B^T e \leq p. \tag{B.17}$$

Let $e = Cz + \mu$. Then this is equivalent to

$$B^T (Cz + \mu) \leq p$$

$$\widetilde{B}^T z \leq p - B^T \mu.$$

where $\widetilde{B} = C^T B$. Let

$$\widetilde{B} = QR$$

be the QR decomposition of $\widetilde{B}$ where R is $(K \times J)$ lower triangular. Since Q is orthogonal $Q^T Q = I$ and $\det(Q) = 1$.

Then defining $z = Qx$, the likelihood conditional on $B(\eta)$ and p can be written

$$f_3 [q, p, B(\eta), \theta] = \int\limits_{R^T x \leq p - B^T \mu} \frac{e^{-0.5 x^T x}}{(2\pi)^{\frac{K}{2}}} dx. \tag{B.18}$$

Start from x_K. Let J_K^+ be the set of rows of C that have positive elements in column K and J_K^- the set with negative elements. Let $D = p - B^T \mu$. Then for all $j \in J_K^+$,

$$-\infty \leq x_K \leq \frac{D_j - \sum_{i<K} R(j,i)\, x_i}{R(j,K)}$$

and for all $j \in J_K^-$,

$$\frac{D_j - \sum_{i<K} R(j,i)\, x_i}{R(j,K)} \leq x_K \leq \infty.$$

So, the bounds on x_K are $x_K \in \left[x_K^L, x_K^H\right]$ where

$$x_K^L = \max\left(-\infty, \max_{j \in J_{d_2}^-}\left(\frac{D_j - \sum_{i<K} R(j,i)\, x_i}{R(j,K)}\right)\right)$$

and

$$x_K^H = \min\left(\infty, \min_{j \in J_K^+}\left(\frac{D_j - \sum_{i<K} R(j,i)\, x_i}{R(j,K)}\right)\right).$$

We repeat the calculation for $j = K-1$ through 1. Then the integral is

$$f_3\left[q, p, B(\eta), \theta\right] = \int_{x_1^L}^{x_1^H} \cdots \int_{x_K^L}^{x_K^H} \frac{e^{-0.5x^T x}}{(2\pi)^{\frac{d_2}{2}}} dx. \tag{B.19}$$

Next for all $j \leq K$ define $u_j = \Phi(x_j)$. Then making the change of variables, the integral is equivalent to

$$f_3\left[q, p, B(\eta), \theta\right] = \int_{u_1^L}^{u_1^H} \cdots \int_{u_K^L}^{u_K^H} du \tag{B.20}$$

where

$$\begin{aligned} u_j^L &= \Phi\left(x_j^L\right) \\ u_j^H &= \Phi\left(x_j^H\right). \end{aligned}$$

Finally, for all $j \leq K$ making the change of variable $u_j = \frac{\left(u_j^H - u_j^L\right)(1+v_j)}{2}$, this is equivalent to

$$f_3\left[q, p, B(\eta), \theta\right] = \int_{-1}^{1} \cdots \int_{-1}^{1} \prod_{j=1}^{K} \left(\frac{u_j^H - u_j^L}{2} \right) dv. \tag{B.21}$$

Integrating out the random coefficients, the Case 3 likelihood is

$$\ln f_3\left(q, p, \theta\right) = \int_{\eta} f_3\left[q, p, B(\eta), \theta\right] \phi(\eta) d\eta.$$

C Data

Tables C.1-C.3 show the most frequently purchased two-item combinations. For completeness, Table C.1 is the same as Table A.3 in the paper.

The tables show the following. While each of the top 5 or 10 two-item combinations has an appreciable market share, in aggregate the top 5 account for only 54.34% of two-item combinations and the top 10 account for only 67.20%. To account for 95% of two-item combinations one must include 105 distinct combinations, which are all the combinations listed in Tables C.1-C.3 below. Most of these combinations have small market shares individually, but together they account for a large share of all two-item baskets. Our model can account for this wide variation in choices of types of fruit, numbers of types chosen, and the quantities of each.

Table C.1: Most frequently purchased 2-item combinations (A)

	Freq.	Pct.	Cum. Pct.
Banana, Apples	101533	25.03	25.03
Banana, Berries+Currants	52141	12.85	37.88
Banana, Easy Peelers	24442	6.03	43.91
Banana, Grapes	23977	5.91	49.82
Apples, Easy Peelers	18363	4.53	54.34
Berries+Currants, Apples	15931	3.93	58.27
Apples, Grapes	12052	2.97	61.24
Berries+Currants, Grapes	8592	2.12	63.36
Avocado, Banana	7915	1.95	65.31
Banana, Pears	7681	1.89	67.20
Apples, Pears	6299	1.55	68.76
Banana, Orange	5746	1.42	70.17
Berries+Currants, Easy Peelers	5506	1.36	71.53
Apples, Orange	5070	1.25	72.78
Easy Peelers, Grapes	4856	1.20	73.98
Banana, Melons	3551	0.88	74.85
Banana, Nectarines	3244	0.80	75.65
Banana, Lemon	3187	0.79	76.44
Banana, Kiwi Fruit	3144	0.78	77.21
Berries+Currants, Cherries	3018	0.74	77.96
Banana, Plums	2916	0.72	78.68
Avocado, Berries+Currants	2514	0.62	79.30
Banana, Cherries	2511	0.62	79.92
Berries+Currants, Melons	2151	0.53	80.45
Berries+Currants, Nectarines	2133	0.53	80.97
Apples, Kiwi Fruit	2043	0.50	81.48
Apples, Lemon	2009	0.50	81.97
Apples, Melons	1898	0.47	82.44
Banana, Grapefruit	1829	0.45	82.89
Apples, Nectarines	1803	0.44	83.33
Apples, Plums	1790	0.44	83.77
Avocado, Apples	1751	0.43	84.21
Grapes, Pears	1745	0.43	84.64
Easy Peelers, Pears	1734	0.43	85.06
Grapes, Orange	1508	0.37	85.44

Note: The table records the frequency with which various 2-item combinations were purchased.

Table C.2: Most frequently purchased 2-item combinations (B)

	Freq.	Pct.	Cum. Pct.
Berries+Currants, Kiwi Fruit	1485	0.37	85.80
Berries+Currants, Orange	1426	0.35	86.15
Banana, Pineapples	1392	0.34	86.50
Berries+Currants, Plums	1391	0.34	86.84
Berries+Currants, Lemon	1285	0.32	87.16
Berries+Currants, Pears	1275	0.31	87.47
Apricot, Banana	1263	0.31	87.78
Grapes, Kiwi Fruit	1262	0.31	88.09
Grapes, Melons	1237	0.30	88.40
Grapes, Plums	1201	0.30	88.69
Banana, Peaches	1126	0.28	88.97
Banana, Mango	1109	0.27	89.24
Easy Peelers, Plums	1087	0.27	89.51
Banana, Dates	1060	0.26	89.77
Easy Peelers, Orange	1060	0.26	90.04
Apples, Grapefruit	986	0.24	90.28
Grapes, Nectarines	980	0.24	90.52
Easy Peelers, Melons	963	0.24	90.76
Easy Peelers, Lemon	949	0.23	90.99
Berries+Currants, Pineapples	899	0.22	91.21
Grapes, Lemon	871	0.21	91.43
Berries+Currants, Peaches	870	0.21	91.64
Easy Peelers, Kiwi Fruit	861	0.21	91.85
Berries+Currants, Mango	842	0.21	92.06
Apples, Pineapples	818	0.20	92.26
Apples, Plums	1790	0.44	83.77
Avocado, Apples	1751	0.43	84.21
Grapes, Pears	1745	0.43	84.64
Easy Peelers, Pears	1734	0.43	85.06
Grapes, Orange	1508	0.37	85.44
Berries+Currants, Kiwi Fruit	1485	0.37	85.80
Berries+Currants, Orange	1426	0.35	86.15
Banana, Pineapples	1392	0.34	86.50
Berries+Currants, Plums	1391	0.34	86.84
Berries+Currants, Lemon	1285	0.32	87.16

Note: The table records the frequency with which various 2-item combinations were purchased.

Table C.3: Most frequently purchased 2-item combinations (C)

	Freq.	Pct.	Cum. Pct.
Berries+Currants, Pears	1275	0.31	87.47
Apricot, Banana	1263	0.31	87.78
Grapes, Kiwi Fruit	1262	0.31	88.09
Grapes, Melons	1237	0.30	88.40
Grapes, Plums	1201	0.30	88.69
Banana, Peaches	1126	0.28	88.97
Banana, Mango	1109	0.27	89.24
Easy Peelers, Plums	1087	0.27	89.51
Banana, Dates	1060	0.26	89.77
Easy Peelers, Orange	1060	0.26	90.04
Apples, Grapefruit	986	0.24	90.28
Grapes, Nectarines	980	0.24	90.52
Easy Peelers, Melons	963	0.24	90.76
Easy Peelers, Lemon	949	0.23	90.99
Berries+Currants, Pineapples	899	0.22	91.21
Grapes, Lemon	871	0.21	91.43
Berries+Currants, Peaches	870	0.21	91.64
Easy Peelers, Kiwi Fruit	861	0.21	91.85
Berries+Currants, Mango	842	0.21	92.06
Apples, Pineapples	818	0.20	92.26
Orange, Pears	818	0.20	92.47
Nectarines, Plums	791	0.19	92.66
Cherries, Apples	774	0.19	92.85
Lemon, Orange	741	0.18	93.03
Avocado, Easy Peelers	699	0.17	93.21
Easy Peelers, Nectarines	691	0.17	93.38
Apricot, Berries+Currants	673	0.17	93.54
Apples, Mango	664	0.16	93.71
Pears, Plums	618	0.15	93.86
Apples, Peaches	611	0.15	94.01
Avocado, Grapes	575	0.14	94.15
Grapes, Pineapples	572	0.14	94.29
Cherries, Grapes	556	0.14	94.43
Lemon, Lime	542	0.13	94.56
Grapes, Grapefruit	513	0.13	94.69

Note: The table records the frequency with which various 2-item combinations were purchased.

Another way to see the variety of choices and the potential role of complementarities is to look at the frequency of basket size conditional on fruit choice. Tables C.4-C.5 show, conditional on purchase of a fruit type, how frequently each basket size was purchased. Except for bananas, cherries, and lemons, all categories are more likely to be purchased in combinations than as stand-alone categories. The relative frequencies of basket size vary across fruit categories and the larger baskets are usually less frequent. These patterns strongly violate the usual independence assumptions of typical discrete choice demand models.

Table C.4: Number of categories purchased conditional on fruit type (A)

	Size of fruit basket						
	1	2	3	4	5	6	Total
Apricot	425	618	656	560	409	681	3349
	12.69	18.45	19.59	16.72	12.21	20.33	100.00
Avocado	5099	4592	3903	2879	1938	2399	20810
	24.50	22.07	18.76	13.83	9.31	11.53	100.00
Banana	121133	103981	71415	39854	20041	15468	371892
	32.57	27.96	19.20	10.72	5.39	4.16	100.00
Berries+Currants	46458	37782	28220	18430	11102	10739	152731
	30.42	24.74	18.48	12.07	7.27	7.03	100.00
Cherries	2611	3296	2778	2040	1336	1731	13792
	18.93	23.90	20.14	14.79	9.69	12.55	100.00
Dates	1104	867	703	494	285	416	3869
	28.53	22.41	18.17	12.77	7.37	10.75	100.00
Apples	59971	76517	59414	34882	18040	14545	263369
	22.77	29.05	22.56	13.24	6.85	5.52	100.00
Easy Peelers	30193	35914	30488	18977	10402	9099	135073
	22.35	26.59	22.57	14.05	7.70	6.74	100.00
Grapes	36085	39580	33187	22622	13088	11627	156189
	23.10	25.34	21.25	14.48	8.38	7.44	100.00
Grapefruit	2387	2985	2930	2567	1857	2522	15248
	15.65	19.58	19.22	16.83	12.18	16.54	100.00
Kiwi Fruit	4297	6561	6821	5705	4081	5062	32527
	13.21	20.17	20.97	17.54	12.55	15.56	100.00
Lemon	8175	7736	6671	5183	3601	4227	35593
	22.97	21.73	18.74	14.56	10.12	11.88	100.00
Lime	975	1372	1302	1082	835	1211	6777
	14.39	20.24	19.21	15.97	12.32	17.87	100.00
Lychees	182	210	226	170	126	200	1114
	16.34	18.85	20.29	15.26	11.31	17.95	100.00

Note: The table records the frequency of each fruit basket size conditional on purchasing the listed fruit category. Column 1 lists the fruit categories. The middle columns record the frequencies. The final column records the total number of observations of each type.

Table C.5: Number of categories purchased conditional on fruit type (B)

	Size of fruit basket						
	1	2	3	4	5	6	Total
Mango	2074	2865	3059	2533	1830	2735	15096
	13.74	18.98	20.26	16.78	12.12	18.12	100.00
Melons	7669	9212	8553	6539	4494	5378	41845
	18.33	22.01	20.44	15.63	10.74	12.85	100.00
Nectarines	6141	8720	8061	6114	4187	4731	37954
	16.18	22.98	21.24	16.11	11.03	12.47	100.00
Orange	12404	15247	13809	9562	5739	5838	62599
	19.82	24.36	22.06	15.28	9.17	9.33	100.00
Passion Fruit	218	317	283	246	200	328	1592
	13.69	19.91	17.78	15.45	12.56	20.60	100.00
Paw-Paws	138	219	234	216	154	261	1222
	11.29	17.92	19.15	17.68	12.60	21.36	100.00
Peaches	2811	3855	3528	2667	1766	2247	16874
	16.66	22.85	20.91	15.81	10.47	13.32	100.00
Pears	11486	20541	22356	16794	10240	9645	91062
	12.61	22.56	24.55	18.44	11.25	10.59	100.00
Pineapples	4857	5352	4905	3959	2734	3675	25482
	19.06	21.00	19.25	15.54	10.73	14.42	100.00
Plums	8947	11592	10874	8150	5423	5893	50879
	17.58	22.78	21.37	16.02	10.66	11.58	100.00
Pomegranates	559	565	454	346	262	288	2474
	22.59	22.84	18.35	13.99	10.59	11.64	100.00
Rhubarb	356	393	380	293	209	236	1867
	19.07	21.05	20.35	15.69	11.19	12.64	100.00
Sharon Fruit	341	375	371	340	266	366	2059
	16.56	18.21	18.02	16.51	12.92	17.78	100.00
Total	377096	401264	325581	213204	124645	121548	1563338
	24.12	25.67	20.83	13.64	7.97	7.77	100.00

Note: The table records the frequency of each fruit basket size conditional on purchasing the listed fruit category. Column 1 lists the fruit categories. The middle columns record the frequencies. The final column records the total number of observations of each type.

D Hedonic price functions

As discussed in Section 6.2 in the paper, for each fruit category we estimate a hedonic price model

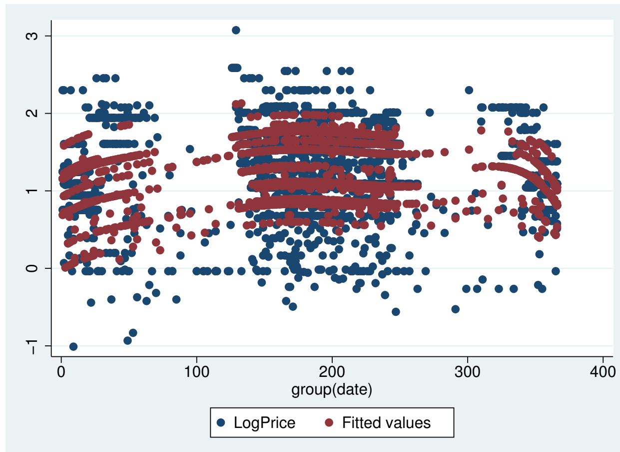
$$\ln p_{it} = \beta x_{it} + h(t) + \varepsilon_{it}$$

where $\ln p_{it}$ is the price of item i in period t, x_{it} is a vector of characteristics of item i in period t and $h(t)$ is a 6th order polynomial function of time. Time is measured as the day within the year. Characteristics included in the regressions are country of origin, branded, organic, tiering (economy, premium or standard), fascia (one of ten firms in the UK or other), packaging, online shop, and small store.

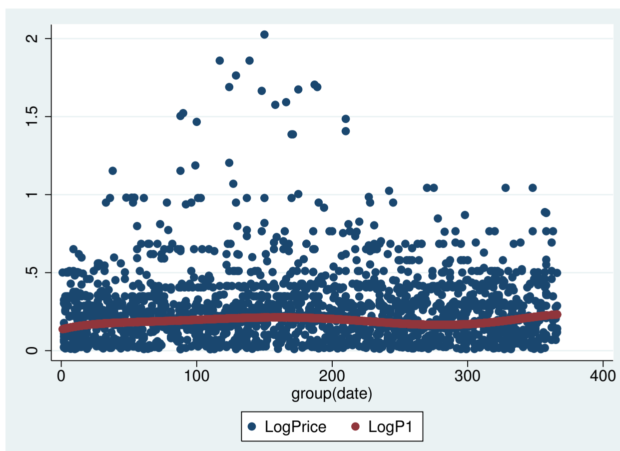
Figure D.1 shows price data and imputed prices for 3 representative examples of the 27 fruit categories: apricots, bananas and cherries. Price is observed for each shopping trip where a particular fruit is purchased. Each figure shows a scatter plot of observed log prices and imputed log prices. For apricots and cherries, prices rise in the spring and the autumn. These are periods when fresh apricots and cherries are more costly and more scarce. In contrast, the price of bananas is relatively flat. The pictures also make clear that at a single point in time there is a great deal of variability in price. This variation is primarily due to variation across fascia and variation due to promotions.

Figure D.1: Prices of apricots, bananas and cherries

(a) Apricots



(b) Bananas



(c) Cherries