Essays on Linear Demand in Industrial Organisation

Afonso Boavida Rodrigues

Keble College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2026

Acknowledgements

I am grateful to my supervisor, Alexei Parakhonyak, for his support throughout my DPhil. Thank you also to Abigail Adams and Mark Armstrong, who for a brief period were also a part of my doctoral journey as supervisors.

For their encouragement, mentorship, and help with various research projects, I am grateful to Howard Smith, Ian Crawford, and Alexander Teytelboym. My work greatly benefited from their deep understanding of their respective fields, their seemingly unlimited availability, and their always timely advice.

Thank you to my colleagues and friends at the Department for their positivity and collegiality, which made my time at Oxford so much more enjoyable. A special thanks to Giulio Gottardo for his close friendship, helpful advice, and reliable company.

Thank you to the Industrial Organisation Reading Group, for enabling me to regularly engage with such a uniquely skilled, curious, and enthusiastic community, and to Eszter Kabos, for her friendship and leadership in helping to bring the group to life.

Thank you to the Department of Economics and the Fundação para a Ciência e Tecnologia, which I gratefully acknowledge for their financial support.

Finally, I am thankful to my sister Leonor, my parents, and my partner Savannah. Without their unwavering and unconditional support, I could not have completed this work. Amo-vos muito.

Contents

1 A Theory of Consumer Choice over Shopping Baskets **1**

1.1 Introduction . . . 2

1.2 Literature Review . . . 4

1.3 Theoretical Framework . . . 7

1.3.1 Setup . . . 7

1.3.2 The Representative Consumer . . . 14

1.3.3 Discussion . . . 20

1.3.4 Equilibrium Analysis . . . 21

1.4 Comparative Statics . . . 24

1.5 Illustrative Example . . . 32

1.6 Discussion and Conclusion . . . 33

2 Empirical Estimation of Aggregated Linear Demand Models **36**

2.1 Introduction . . . 36

2.2 Literature Review . . . 40

2.3 Data . . . 45

2.4 Empirical Application . . . 50

2.4.1 Empirical Specification . . . 50

2.4.2 Identification Strategy . . . 59

2.5 Results . . . 60

2.6 Discussion and Conclusion . . . 72

3 Endogenous Product Design in Aggregated Linear Demand Models **74**

3.1 Introduction . . . 75

3.2 Literature Review . . . 78

3.3 Theoretical Framework . . . 82

3.3.1 Setup . . . 82

3.3.2 Characteristics-Based Linear Demand . . . 83

3.4 Optimal Product Design under Monopoly Pricing . . . 89

3.4.1 One-Attribute Case . . . 89

3.4.2 Attribute Exclusivity . . . 91

3.4.3 Attribute Non-Exclusivity . . . 93

3.4.4 Discussion . . . 98
3.5 Optimal Product Design under Single-Product Firm Pricing . . . 99
3.5.1 Attribute Exclusivity . . . 100
3.5.2 Attribute Non-Exclusivity . . . 102
3.5.3 Discussion . . . 108
3.6 Firm- and Product-level Heterogeneity in Attribute Costs . . . 110
3.7 Product Entry and Market Outcomes . . . 115
3.8 Numerical Example . . . 118
3.9 Discussion and Conclusion . . . 120

Appendices

A Appendix - Chapter 1 **131**
A.1 Proofs . . . 131
A.2 Welfare analysis . . . 134
A.3 Details on usage of transaction sample data . . . 136
A.4 Accounting for multi-product monopolistic competition . . . 136
A.5 Retail stock-outs . . . 137

B Appendix - Chapter 2 **140**
B.1 Handling less-than-full-row-rank consideration sets in empirical estimation 141
B.2 Participation rates . . . 142
B.2.1 Transaction trends . . . 144
B.2.2 Estimation of participation rate elasticity . . . 145

C Appendix - Chapter 3 **147**
C.1 Proofs . . . 147
C.2 Finding S when X is not observed but M is . . . 150
C.3 Connection between M and $\mathbb{E}(\boldsymbol{\delta}\boldsymbol{\delta}')$ . . . 151
C.4 Exploratory analysis of empirical testability . . . 151
C.4.1 Estimating U . . . 152
C.4.2 Estimating Γ . . . 156

List of Figures

1.1 Example of shopping baskets in a consumer's consideration set 10

2.1 Shop locations 46
2.2 Revenue-weighted mean prices 49
2.3 Proxy heatmap of own- and cross-category substitution - $W^{(s)}$ 56
2.4 Proxy heatmap of own- and cross-category complementarity - $W^{(c)}$ . . . 56
2.5 Revenue-weighted mean mark-ups - product- and retailer-level 64
2.6 Revenue-weighted four-quarter moving average mark-ups: U.S. survey estimate and sample 65
2.7 Average mark-up distribution - selected categories 66
2.8 Revenue-weighted mean mark-ups - percentiles 67
2.9 Revenue-weighted mean mark-ups - supermarkets 68
2.10 Mark-up transition dynamics and transition shares over time 69
2.11 Fixed-basket mark-up indices 70

3.1 Selected sample attribute choice simulations under symmetric costs . . . 110
3.2 Selected sample attribute choice simulations under asymmetric costs . . . 112

B.1 Per-supermarket transaction trends 144

List of Tables

1.1 Simulation of consumer choice counterfactuals 28
1.2 Welfare outcomes under various market structure and consideration set constraint specifications. 33
1.3 Average market outcomes under various market structure and consideration set constraint specifications. 33

2.1 Transaction statistics across products and categories 46
2.2 Transaction statistics per store 47
2.3 Product availability in stores and the average revenue brought by products at each level of availability 47
2.4 Shop statistics 47
2.5 Product category statistics 48
2.6 Proxy details - top categories by cross-category exposure 55
2.7 Proxy details - top categories by own-category exposure 55
2.8 Standardised 2SLS first stage 61
2.9 Model diagnostics and PCA summary 62
2.10 Standardised 2SLS parameters of interest (Pre-NNLS vs Post-NNLS) . . 63
2.11 Revenue-weighted median own-price elasticity by category 71

B.1 Participation shares - IV results by specification 146

Abstract

Demand estimation refers to the practice of recovering consumers' demand for products from data. It helps determine how much consumers value different product characteristics, how sensitive demand is to price, how substitutable products are, and how consumers' choices change under counterfactuals such as mergers, entry, taxes, or price changes.

Characteristics-based discrete-choice models are the tool most commonly used for estimating demand in differentiated products markets. Consumers are assumed to select one option from a finite choice set, based on the options' observable characteristics and their own preferences over those characteristics. These models are especially capable in settings with substantial consumer heterogeneity and sufficient data availability. However, they are often inflexible and computationally intractable. Especially problematic are large sets of alternatives, multi-product transactions, and complementarity, cross-category or otherwise; model outcomes are also often sensitive to the choice of market boundaries. This has had an impact on the scope of research questions which economists are able to address. At times, it has raised issues for antitrust litigation. Matters of interest to regulators, consumers, and firms alike have gone under explored. These include: how demand and supply shocks propagate through large product assortments; how excise taxes affect staple good demand, or to what degree private labels change market power dynamics. It has also been suggested that market boundaries should be informed by estimated substitution and complementarity effects, rather than set exogenously. Discrete-choice models do not address these issues convincingly.

Continuous demand systems have been suggested in the past as an alternative for discrete-choice models. They treat each product as a distinct source of consumer utility and specify the quantity demanded for each as a continuous function of the full price vector, rather than modeling any single discrete choice among alternatives. *Linear demand* is the most well-known continuous demand system. Due to its flexibility, (computational) tractability, and scalability, it is especially common in international trade and microeconomic theory. It has its own major issues, however, which has prevented serious usage in empirical microeconomics, and more specifically industrial organisation.

In this thesis, I aim to address the major concerns raised in the literature regarding linear demand. My research goal is to encourage its wider use in microeconomic research, broadening the scope of questions economists consider.

In Chapter 1, titled "A Theory of Consumer Choice over Shopping Baskets", I introduce a novel theoretical approach for analysing consumer choice in differentiated products markets with very large choice sets, and where purchases over multiple goods, multiple units, and across product categories are common. Rational inattention theory suggests that consumers tend to restrict their choices to a subset of available alternatives called the *consideration set*. I propose that these consideration sets may include both individual and combinations of goods, which I call *shopping baskets*. Consumers maximise utility through their choice of how much of each of these shopping baskets to purchase. A consumer's consideration set therefore binds their effective consumption bundle. First, I show that the model is tractable and well-behaved, and delivers unique linear demand. Second, I demonstrate how consideration sets can help microfound a representative consumer over aggregate linear demand from consumers with heterogeneous preferences. Microfoundations are necessary where only market-level data is available. Demand estimation over aggregated data can then tie model parameters to underlying consumer population primitives. This constitutes the first step towards overcoming concerns regarding linear demand: existing microfoundations require unrealistic consumer homogeneity. I provide a full characterisation of equilibrium behaviour between said representative consumer and price-setting firms across different market structures, and discuss the implications of consideration-set-constrained consumer choice for market outcomes.

Chapter 2, titled "Empirical Estimation of Aggregated Linear Demand Models", applies Chapter 1's approach. I use anonymised transaction-level data for an empirical study of intra-store competition in the Portuguese supermarket industry between 2020 and 2023. Doing so offers a unique window into how retail dynamics shift under significant economic pressure and rapidly evolving consumer habits. Here, linear demand estimation finds three other well-known limitations. In depending on the full price vector, linear demand suffers from a *curse of dimensionality.* Since as the number of product grows, the number of price parameters grows quadratically, data becomes too sparse in the high-dimensional space to estimate demand precisely without unreasonably large samples. To overcome price endogeneity, every price must also be instrumented (non-collinearly) for each demand function. Lastly, estimating aggregate demand requires adjusting for aggregated individual and heterogeneous corner solutions in shopping baskets choice, which the representative consumer rationalisation above highlights. I show how consideration sets can once again be leveraged to overcome all thee issues. I demonstrate the model by estimating price elasticities and assessing market power across almost 30000 goods and more than 500 product categories.

Chapter 3, titled "Endogenous Product Design in Aggregated Linear Demand Models", introduces a novel characteristics-based specification for linear demand. Characteristics are allowed to affect both consumers' product valuations and to what extent these compete. This Chapter demonstrates how such a specification can help linear demand deliver fresh insights in how firms may optimally design existing goods, as well as predict demand for new products. I show the model is tractable enough to solve settings with any finite number of goods, firms, and product characteristics, with both vertical and horizontal differentiation across different market structures, and under firm asymmetry.

1

A Theory of Consumer Choice over Shopping Baskets

Contents

1.1 Introduction **2**

1.2 Literature Review **4**

1.3 Theoretical Framework **7**

1.3.1 Setup 7

1.3.2 The Representative Consumer 14

1.3.3 Discussion 20

1.3.4 Equilibrium Analysis 21

1.4 Comparative Statics **24**

1.5 Illustrative Example **32**

1.6 Discussion and Conclusion **33**

1.1 Introduction

In this Chapter, I introduce a novel approach to modelling continuous demand systems, utilising consideration sets to analyse differentiated products markets with very large choice sets, and where purchases over multiple goods, multiple units, and across product categories are common.[1] These types of transactions are ubiquitous in everyday economic activity, be it during grocery shopping, dining at a restaurant, or making investment portfolio decisions. However, we still know very little of the ways in which they differ from other kinds of purchases, and what they imply for aggregate demand and other market-level outcomes.

First, I propose a theory of consumer choice, whereby consumers' individual consumption bundles are constrained by the choices they make over finite, latent, stable consideration sets. These consideration sets are made up of different unique combinations of goods, which I call *shopping baskets*. Consumers maximise utility through their choice of shopping baskets, meaning their effective consumption bundle is constrained to the vector space enabled by the consideration set. I accommodate broad flexibility in consumer choice while imposing realistic constraints. I work throughout with a quasilinear quadratic utility specification, as a second-order local approximation to a general, smooth utility function. This delivers a linear demand system that is simple enough for high-dimensional theoretical (but also empirical) work.

Second, I provide an aggregation result, whereby aggregated demand across a population of consumers with heterogeneous preferences and consideration sets is shown to be

[1] I have no relevant or material financial interests that relate to the research described in this Chapter. I gratefully acknowledge financial support by the Department of Economics at the University of Oxford and the Fundação para a Ciência e Tecnologia of the Ministry of Education, Science and Innovation (Portugal). This Chapter has benefited from thoughtful conference discussions by Leonardo Madio, Anne-Christine Barthel, and Alminas Zaldokas, as well as valuable comments from participants of the NIE PhD Symposium 2024, the 13th Oligo Workshop, the Research Jamboree at the University of Oxford, the Lisbon Meetings in Game Theory and Applications #14, EARIE 2025, the CEPR Paris Symposium 2025, and MaCCI 2026.

rationalisable as the solution of a representative consumer's constrained maximisation problem in the same local functional class, over a single aggregated consideration set. It implies that the model remains empirically and theoretically useful even when only market-level data is observed, and provides conditions under which underlying population preference parameters can be learned. Extending Gorman (1953, 1961), which admits a representative consumer rationalisation only under restrictive consumer homogeneity, this result enables deep heterogeneity in preferences, price sensitivity, and consideration sets. The only conditions for the result to hold is (i) that local demand slopes be proportional across all consumers, and (ii) that a sufficiently rich aggregate consideration set be observed, obtained from a combination of unique shopping baskets across all consumers' consideration sets. The latter is later shown to hold in general for even just a small number of consumers, while the former corresponds to an assumption that there be an approximate agreement amongst consumers as to which goods are or not substitutes.

Lastly, I provide a full characterisation of equilibrium behaviour between said representative consumer and price-setting firms across different market structures, and discuss the implications of consideration-set-constrained consumer choice for market outcomes. I show that if the consideration set binds, joint purchases induce substitutes and complements onto the system. When the structure of the consideration set binds, own-price effects are weaker than they would otherwise be, as goods' demand becomes relatively more dependent on those goods with which they are often co-purchased. Often co-considered goods behave as complements, while goods often paired with the same products face substitution effects conditional on these. The constrained choice behaviour imposes lower average demand, aggregate profits, and consumer surplus. A major implication of the outsized role of consideration is that product ownership concentration may become welfare-enhancing.

In short, this Chapter reveals the important role that consideration sets have for understanding how individual choice over combinations of goods affects aggregate outcomes.

The rest of this Chapter is structured as follows. Section 1.2 reviews the relevant literature. Section 1.3 introduces the theoretical framework, beginning with the model setup, then examining aggregation through a representative agent, and finally characterising equilibrium outcomes under different market structures. Section 1.4 studies the comparative statics of the model and complements the theoretical results with Monte Carlo evidence. Section 1.5 illustrates the main mechanisms through a simple example. Section 1.6 concludes.

1.2 Literature Review

This Chapter contributes to three lines of inquiry in empirical microeconomics research.

Relevant to the consumer choice literature, I pursue a framework based on consideration sets for thinking about multi-unit, multi-good, cross-category purchases. This differs from the multi-level demand systems of Gorman (1953, 1961), Deaton and Muellbauer (1980), and Hausman, Leonard, and Zona (1994), where the weak separability assumption in the utility function blocks most cross-group interactions, except via the top budget. I adopt a tractable linear-demand framework, derived from quasi-linear quadratic utility (see Amir, Erickson, and Jin, 2017, for a deep dive; see Choné and Linnemer, 2020, for a historical retrospective). Linear demand remains the workhorse model in Economics textbooks due to straightforward intuition and application, closed-form solutions, and parsimony. Quasi-linear quadratic utility, from whence linear demand is usually derived, is also a local second-order approximation to any utility function, allowing a generalisability of results across different utility functional specifications and underlying consumer behaviour.[2]

[2]Linear demand is itself a local first-order approximation to any well-behaved demand function.

Much has been said about the ability of representative consumer models to approximate aggregate behaviour (see e.g. Anderson, de Palma, and Thisse, 1992). While representative consumer models with a quasilinear quadratic utility/linear demand specification are popular in economic theory (e.g. Ballester, Calvó-Armengol and Zenou, 2006), international trade (e.g. Melitz and Ottaviano, 2008), and finance (e.g. Campbell and Viceira, 2002, Ch.2), in empirical microeconomics, more specifically in industrial organisation, their use has remained limited. In the literature discussing representative consumer approaches, the most common concern lies with approximating the individual corner solutions of a heterogeneous population of consumers, as consumers with different preferences will buy only some of the variants on offer, and even amongst those most commonly bought they will have different prices at which consumption is not longer desirable. Gorman (1953) provides a well-known result in this literature, whereby, under sufficient consumer homogeneity, aggregation over linear demands is possible. This is said to be a highly restrictive condition - Kirman (1992), Carroll (2000) and Cherchye, Crawford, De Rock, and Vermeulen (2016) more recently all find strong evidence against the existence of a representative consumer under that standard. One of the major contributions of this Chapter is a novel aggregation result, whereby consumers sharing proportional demand slopes across goods can otherwise be fully heterogeneous in preferences and consideration sets while remaining rationalisable by a representative agent maximising the same utility functional form over the aggregate consideration set derived from those of the underlying consumer population. This demand slope assumption, seemingly restrictive, is in fact along the lines of Eaton and Lipsey (1989)'s "awkward facts" on product differentiation, namely that while consumers perceive differences between goods within a market, there is generally an approximate agreement as to which ones are or not substitutes (examples given there include low- versus high-tar cigarrettes and compact versus subcompact cars).

Consideration sets themselves satisfy another of these "awkward facts": that consumers purchase only a small subset of those products available in a given market. Consideration sets, first established in Wright and Barbour (1977), are a relaxation of the assumption that individuals consider all feasible choices. They can be defined as the restricted set of shopping baskets to which a consumer pays attention in their choice process. Many reasons have been given as to why the choice set (the set of available choices) may sometimes differ from a consumer's consideration set: behavioural heuristics (Hauser, 2014), cognitive limitations (Simon, 1957), bounded rationality and seller persuasion (Eliaz and Spiegler, 2011), imperfect information (Sovinsky, 2008), search (Caplin, Dean, and Martin, 2011), or evaluation costs broadly speaking (Hauser and Wernerfelt, 1990). One concern the present Chapter speaks to in this literature is the adjustment of consideration sets to settings where it is probable that consumers consider goods both individually and in combination (e.g. Manchanda, Ansari, and Gupta, 1999; Fox and Lazzati, 2017; Allen and Rehbeck, 2022).

Nocke and Schutz (2025) is closely related to this Chapter's approach. The paper takes a potential-games approach, introducing transformed potentials and characterising microfounded IIA/generalised-linear and nest/basket demand classes that permit complementarities and guarantee equilibrium existence under multi-product firm-pricing games. Nocke and Schutz (2025) provides a complementary theoretical characterisation to the present Chapter, though one with a similarly unrestricted consideration set. Nocke and Schutz (2025) find that whether products are substitutes or complements depends on the interaction effects between goods and outside option but also through joint purchases - the extent to which the products of two different firms are predominantly purchased together or not determines the strength of complementarity. The present Chapter complements this work by extending representative consumer microfoundations to heterogeneous consideration sets, while providing an empirical scalability (discussed in Chapter 2).

1.3 Theoretical Framework

1.3.1 Setup

Let there be N consumers, indexed $i = 1, \ldots, N$, and a finite number of non-numeraire goods K, indexed $k = 1, \ldots, K$. A consumption bundle $\boldsymbol{q}_i \in \mathbb{R}^K$ denotes the quantities consumed by i of each of the K goods. Let $q_{i0} \in \mathbb{R}$ denote consumption of a numeraire good whose price is normalised to one. Assume that consumers hold continuous and monotonic rational preferences such that these admit a continuous utility function $U_i(\boldsymbol{q}_i, q_0)$, $\forall i$, over alternative choices of $(\boldsymbol{q}_i, q_{i0})$. We assume this utility function to be twice-differentiable, and admit a quasilinear representation of the form:

$$U_i(\boldsymbol{q}_i, q_{i0}) = U_i(\boldsymbol{q}_i) - \phi_i q_{i0} \tag{1.1}$$

for $-\phi > 0$ the marginal utility of income and $U_i(\boldsymbol{q}_i)$ the sub-utility obtained from the K non-numeraire goods. Consumer i has income/wealth Y_i. They spend part of it on the K non-numeraire goods, and the remaining on the numeraire good. Normalising the numeraire price to one, the quantity of the numeraire equals the amount of money left over, i.e. the following budget constraint:

$$q_{i0} = Y_i - \boldsymbol{q}_i'\boldsymbol{p} \tag{1.2}$$

for $\boldsymbol{p}$ the price vector for the non-numeraire goods. The quasilinear specification implies that the numeraire enters utility linearly and separably. Consequently, all curvature in preferences is captured by $U_i(\boldsymbol{q}_i)$, while the marginal utility of income $-\phi$ is constant.

For any such sub-utility function $U_i(\boldsymbol{q}_i)$, we can obtain a quadratic representation via second-order Taylor expansion around a reference bundle ($\overline{\boldsymbol{q}}$):

$$
\begin{aligned}
U_i(\boldsymbol{q}_i) &\approx U_i(\overline{\boldsymbol{q}}) + (\boldsymbol{q}_i - \overline{\boldsymbol{q}})'\nabla U_i(\overline{\boldsymbol{q}}) + \frac{1}{2}(\boldsymbol{q}_i - \overline{\boldsymbol{q}})'\nabla^2 U_i(\overline{\boldsymbol{q}})(\boldsymbol{q}_i - \overline{\boldsymbol{q}}) \\
&\approx \boldsymbol{q}_i' \underbrace{(\nabla U_i(\overline{\boldsymbol{q}}) - (\nabla^2 U_i(\overline{\boldsymbol{q}}))\overline{\boldsymbol{q}})}_{\boldsymbol{\delta}_i,\text{ marginal util. at reference bundle}} + \frac{1}{2}\boldsymbol{q}_i' \underbrace{(\nabla^2 U_i(\overline{\boldsymbol{q}}))}_{-M_i,\text{ Hessian}} \boldsymbol{q}_i \\
&\quad + \underbrace{[U_i(\overline{\boldsymbol{q}}) - \nabla U_i(\overline{\boldsymbol{q}})'\overline{\boldsymbol{q}} + \frac{1}{2}\overline{\boldsymbol{q}}'(\nabla^2 U_i(\overline{\boldsymbol{q}}))\overline{\boldsymbol{q}}]}_{u_{i0},\text{ utility level at reference bundle}} \\
&\approx \boldsymbol{q}_i'\boldsymbol{\delta}_i - \frac{1}{2}\boldsymbol{q}_i' M_i \boldsymbol{q}_i + u_{i0}
\end{aligned}
\tag{1.3}
$$

At $\overline{\boldsymbol{q}} = 0$, $u_{i0} = U_i(0)$; we set this to zero across all i without loss of generality (WLOG) for our utility maximisation exercise below. At $\overline{\boldsymbol{q}} = 0$, we also have $\boldsymbol{\delta}_i = \nabla U_i(0)$, which can then be reinterpreted as the initial marginal utility per good for consumer i. The indirect utility function of said consumer is then:

$$
\begin{aligned}
v_i(\boldsymbol{p}, Y_i) &= -\phi_i Y_i + \max_{\boldsymbol{q}_i}(U_i(\boldsymbol{q}_i) + \phi_i \boldsymbol{q}_i'\boldsymbol{p}) \\
&= -\phi_i Y_i + \frac{1}{2}(\boldsymbol{\delta}_i + \phi_i\boldsymbol{p})' M_i^{-1}(\boldsymbol{\delta}_i + \phi_i \boldsymbol{p})
\end{aligned}
\tag{1.4}
$$

Applying Roy's identity:

$$
\boldsymbol{q}_i(\boldsymbol{p}, Y_i) = -\frac{\nabla_{\boldsymbol{p}} v_i(\boldsymbol{p}, Y_i)}{\partial v_i(\boldsymbol{p}, Y_i)/\partial Y_i} = M_i^{-1}(\boldsymbol{\delta}_i + \phi_i \boldsymbol{p}) \tag{1.5}
$$

In other words, quasilinear quadratic utility implies an exact linear demand, with zero income effects for the non-numeraire goods. Throughout this Chapter, I assume M_i is symmetric and positive definite for all i. This guarantees the invertibility of M_i and is necessary and sufficient for the well-behavedness of $U_i(\boldsymbol{q}_i)$, $\forall i$, such that it admits a unique global maximum in $\boldsymbol{q}_i$ (Amir et al., 2017). Quasilinearity implies no income effects, a common simplification particularly in response to local changes in prices (Choné and Linnemer, 2020), and for goods individually representing a small portion of total income (Vives, 1987).

A common assumption is to let, as we have so far, $\boldsymbol{q}_i \in \mathbb{R}^K$ - i.e. the consumption bundle can be any vector of K non-negative real numbers. However, for K goods, there are up to 2^K different feasible consumption bundles; for sufficiently large K, this assumption is therefore too strict. In this paper, I instead assume rational inattention; physical and technological conditions, mental capacity, and/or individual preferences are said to constrain the space of relevant bundles. In line with the literature, I call the subset of all feasible alternative consumption bundles that a consumer pays attention to and evaluates when making a decision the consumer's *consideration set.*

Let $\mathcal{J}_i$ be the discrete set of these alternatives, or *shopping baskets*, considered - individually or in linear combination - by consumer i and indexed $j = 1, \ldots, J_i$; and $A_{K \times J_i}$, the matrix representation of the consideration set: a sparse matrix of product presence per unique basket considered, each element a_{kj} constituting the number of units of a given good k in a given basket j considered.[3] The columns of matrix A_i are non-zero and unique, though not necessarily linearly independent.

Consumers decide which and how many alternatives to purchase. To account for the potential indivisibility of said alternatives, we can model the intensive and extensive margins of this choice via $\hat{\boldsymbol{z}}_{J_i \times 1}$, the unit-normalised choice bundle ($\mathbf{1}'\hat{\boldsymbol{z}} = 1$); and T_i, the scalar number of transactions pursued. Both can be directly observed from transaction data.

[3]Each column can be thought of as the equivalent to one possible receipt during a purchase instance.

Figure 1.1: Example of shopping baskets in a consumer's consideration set

```
ID: C236-491
01/01/2026
--------
ITEM      QTY
Milk        2
Bread       1
Apples      6
Rice        3
 END RECEIPT
```

```
ID: C431-667
02/01/2026
--------
ITEM      QTY
Milk        1
Bread       2
Banana      4
Rice        1
 END RECEIPT
```

$$\Rightarrow \qquad A_i = \begin{bmatrix} \cdot & 2 & 1 & \cdot \\ \cdot & 1 & 2 & \cdot \\ \cdot & 6 & 0 & \cdot \\ \cdot & 0 & 4 & \cdot \\ \cdot & 3 & 1 & \cdot \end{bmatrix}$$

Notes: Each consumer's consideration set incorporates all alternative shopping baskets considered by the consumer. Receipts of purchases made by a consumer are one way to identify the structure of said consumer's consideration set.

We may thus express the goods consumption bundle $\boldsymbol{q}_i$ as a *non-negative linear* function of a *normalised* $\hat{\boldsymbol{z}}_i$:

$$\boldsymbol{q}_i = T_i A_i \hat{\boldsymbol{z}}_i \qquad \& \qquad T_i, \hat{\boldsymbol{z}}_i \geq 0 \qquad \& \qquad \mathbf{1}'\hat{\boldsymbol{z}}_i = 1 \tag{NN-LF-NORM}$$

The NN-LF-NORM condition summarises the constraints imposed by a consumer's consideration set on choice. When a consumer i makes purchase decisions, the set of all feasible goods consumption bundles will span the *conical hull* of matrix A_i:

$$\begin{aligned} &\boldsymbol{q}_i/T_i \in \operatorname{conv}(A_i) = \Big\{ \sum_{j=1}^{J} \boldsymbol{a}_{ij} x_j : \boldsymbol{a}_{ij} = A_i \boldsymbol{e}_j, x_j \in \mathbb{R}_+^K, \mathbf{1}'\boldsymbol{x} = 1, j \in \mathcal{J}_i \Big\} \\ \Leftrightarrow\ &\boldsymbol{q}_i \in \bigcup_{T_i \geq 0} T_i \cdot \operatorname{conv}(A_i) = \Big\{ \sum_{j=1}^{J} \boldsymbol{a}_{ij} x_j : \boldsymbol{a}_{ij} = A_i \boldsymbol{e}_j, x_j \in \mathbb{R}_+^K, j \in \mathcal{J} \Big\} \\ = &\operatorname{cone}(A_i) \end{aligned} \tag{1.6}$$

Vector $\boldsymbol{e}_j$ is the j-th standard basis vector. Demand for goods per transaction depends on the proportion of choices that favour each shopping basket considered by the consumer. For $T_i \geq 0$, one can scale those proportions up and down, as to get a cone. It is then trivial to see that whether separately optimising over the intensive and extensive margins or doing so jointly makes no difference for the choice of the optimal consumption bundle:

no constraint is imposed.

Instead, the NN-LF-NORM condition will play a role depending on the answer to the following question: when is $\text{cone}(A_i) \subsetneq \mathbb{R}_+^K$?

Proposition 1.1: *If A_i is less than full row rank, $\boldsymbol{q}_i \in \text{cone}(A_i) \subsetneq \mathbb{R}_+^K$. If and only if a standard basis vector $\boldsymbol{e}_i \notin \text{cone}(A_i)$, then $\boldsymbol{q}_i \in \text{cone}(A_i) \subsetneq \mathbb{R}_+^K$.*

Proof: The first part of **Proposition 1.1** follows directly from linear dependence in the rows of a less-than-full-row-rank matrix, and helps us isolate the linear-function (LF) condition from the broader NN-LF-NORM. The second part of the proposition can be proved as follows. Suppose $\text{cone}(A_i) \subsetneq \mathbb{R}_+^K$. Then there exists at least one vector $\boldsymbol{x}_i \in \mathbb{R}_+^K$ such that $\boldsymbol{x}_i \notin \text{cone}(A_i)$. In particular, consider the standard basis vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_K$. If all of these vectors were in $\text{cone}(A_i)$, then by taking nonnegative linear combinations, we could generate any vector in $\mathbb{R}_+^K$, implying $\text{cone}(A_i) = \mathbb{R}_+^K$, a contradiction; $\text{cone}(A_i) \subsetneq \mathbb{R}_+^K$ implies that at least one $\boldsymbol{e}_k \notin \text{cone}(A_i)$. Conversely, suppose there exists some k such that $\boldsymbol{e}_k \notin \text{cone}(A_i)$. Then, the cone cannot generate all vectors in $\mathbb{R}_+^K$, because it fails to generate the direction corresponding to good k alone. Thus, $\text{cone}(A_i)$ must be a strict subset: $\text{cone}(A_i) \subsetneq \mathbb{R}_+^K$. □

In other words, the NN-LF-NORM condition holds if and only if not all goods are considered as standalone alternatives by a consumer.

For the remainder of this Section, I study how this condition affects individual demand functions and verify if they are well-behaved. We can write the consumer objective function's Lagrangean as:

$$\mathcal{L}(T_i, \hat{\boldsymbol{z}}_i) = T_i \hat{\boldsymbol{z}}_i' A_i' \boldsymbol{\delta}_i - \frac{1}{2} T_i^2 \hat{\boldsymbol{z}}_i' (A_i' M_i A_i) \hat{\boldsymbol{z}}_i - \phi_i \Big(Y_i - T_i \boldsymbol{p}' A_i \hat{\boldsymbol{z}}_i \Big) + \underbrace{\hat{\boldsymbol{\lambda}}_i'}_{\text{non-negativity multiplier}} \hat{\boldsymbol{z}}_i + \underbrace{\eta_i}_{\text{unit-normalisation multiplier}} \mathbf{1}' \hat{\boldsymbol{z}}_i \tag{1.7}$$

Maximising the Lagrangian of the consumer's basket choice problem in $\hat{\boldsymbol{z}}_i$, we obtain the following Karush-Kuhn-Tucker (KKT) conditions:

$$T_i (A_i' M_i A_i) \hat{\boldsymbol{z}}_i = A_i'(\boldsymbol{\delta} + \phi_i \boldsymbol{p}) + \frac{1}{T_i}(\hat{\boldsymbol{\lambda}}_{\boldsymbol{i}} + \eta_i \mathbf{1}) \quad \hat{\boldsymbol{\lambda}}_i \geq 0 \quad \hat{\boldsymbol{z}}_i \geq 0 \qquad \hat{\boldsymbol{\lambda}} \circ \hat{\boldsymbol{z}}_i = \mathbf{0} \tag{1.8}$$

Expressing the KKT stationarity condition in terms of $\hat{\boldsymbol{z}}_i$ introduces the first problem derived from our dimensionality constraint.

***Lemma 1.1**: $A_i' M_i A_i$ is positive semi-definite, and therefore not generally invertible.*[4]

To proceed, I multiply both sides of said condition by $(A_i' M_i A_i)^+$, for $(\cdot)^+$ the Moore-Penrose pseudoinverse (Moore, 1920; Penrose, 1955). However, on the LHS, $(A_i' M_i A_i)^+ (A_i' M_i A_i) \neq I$ does not hold in general. A general demand function relative to $\hat{\boldsymbol{z}}_i$ must account for the so-called *homogeneous part* of differential equations:

$$T_i \hat{\boldsymbol{z}}_i = (A_i' M_i A_i)^+ \Big[A_i'(\boldsymbol{\delta}_{\boldsymbol{i}} + \phi_i \boldsymbol{p}) + \frac{1}{T_i}(\hat{\boldsymbol{\lambda}}_{\boldsymbol{i}} + \eta_i \mathbf{1}) \Big] + (I - (A_i' M_i A_i)^+ (A_i' M_i A_i)) \boldsymbol{y} \tag{1.9}$$

for any $\boldsymbol{y} \in \mathbb{R}^J$.[5] There are thus a multiplicity of mappings of $\boldsymbol{p}$, as the strict concavity condition we had imposed on M_i no longer holds for $A_i' M_i A_i$. In other words, the utility-maximising $\hat{\boldsymbol{z}}_i(\boldsymbol{p})$ is non-unique. Nonetheless,

---

[4] Proofs of all Lemmas may be found in the Appendix.

[5] One may verify the propriety of this operation by noting that $(A_i' M_i A_i)(I - (A_i' M_i A_i)^+ (A_i' M_i A_i)) = 0$, and $(A_i' M_i A_i)(A_i' M_i A_i)^+ A_i'(\boldsymbol{\delta}_{\boldsymbol{i}} + \phi_i \boldsymbol{p}) = A_i'(\boldsymbol{\delta}_{\boldsymbol{i}} + \phi_i \boldsymbol{p})$. The latter follows from the properties of the pseudo-inverse, as both $A_i' M_i A_i$ and $A_i'(\boldsymbol{\delta}_{\boldsymbol{i}} + \phi_i \boldsymbol{p})$ are in the column space of A_i'.

Proposition 1.2: *For the set of non-unique $\hat{\boldsymbol{z}}_i$, there is a unique $\boldsymbol{q}_i(\boldsymbol{p}) = T_i A_i \hat{\boldsymbol{z}}_i$:*

$$\boldsymbol{q}_i(\boldsymbol{p}) = A_i(A_i' M_i A_i)^+ \Big[A_i'(\boldsymbol{\delta_i} + \phi_i \boldsymbol{p}) + \frac{1}{T_i}\hat{\boldsymbol{\lambda}}_{\boldsymbol{i}}\Big] \tag{1.10}$$

Proof: Multiplying both sides by A_i, $\boldsymbol{q_i} = T_i A_i \hat{\boldsymbol{z}}_i$ and $A_i(I - (A_i' M_i A_i)^+ (A_i' M_i A_i))\boldsymbol{y} = 0$, as $I - (A_i' M_i A_i)^+ (A_i' M_i A_i)$ is in the null space of $A_i' M_i A_i$, which itself matches the null space of A_i. A_i and $A_i' M A_i$ share a null space as: if $A_i \boldsymbol{y} = 0$, then $A_i' M_i A_i \boldsymbol{y} = 0$ trivially; $\boldsymbol{y}' A_i' M_i A_i \boldsymbol{y} = (A_i \boldsymbol{y})' M_i (A_i \boldsymbol{y}) = 0$ if $A\boldsymbol{y} = 0$. Since M_i is assumed positive definite, the only way for us to have $(A_i \boldsymbol{y})' M_i (A_i \boldsymbol{y}) = 0$ is precisely if $A\boldsymbol{y} = 0$. Last, η is shown to not constrain $\boldsymbol{q}_i$ - i.e. $\eta = 0$. This is the result of the following:

Lemma 1.2: *For an individual consumer and shopping basket choice vector $\hat{\boldsymbol{z}}_i$, the unit-normalisation multiplier η does not constrain $\boldsymbol{q}_i$ - i.e. $\eta = 0$.* □

In short: despite the NN-LF-NORM condition breaking the strict concavity assumption generally required for the well-behavedness of the utility function, its ensuing use (following the pseudoinverse and resulting non-unique $\hat{\boldsymbol{z}}_i$) ultimately isolates the single unique mapping from $\boldsymbol{p}$ to $\boldsymbol{q}_i$. The utility-maximising $\boldsymbol{q}_i(\boldsymbol{p})$ is unique, even if $\hat{\boldsymbol{z}}_i(\boldsymbol{p})$ is not. The above also shows that we need not worry about the NORM constraint, as it does not constrain at all; we can assume divisibility in $\hat{\boldsymbol{z}}_i$ WLOG, as allowing for separate optimisation of the intensive and extensive margin of shopping basket choice does not change optimal $\boldsymbol{q}_i$. Separately considering the number of transactions pursued per consumer may nonetheless be relevant to demand aggregation, which we will get to later.

For S_i the latent unconstrained Slutsky matrix of consumer i's demand function - which is to say, that which we obtain where the Lagrangean constraints do not bind - we have that:

$$S_{i,ab} = \frac{\partial q_{ia}}{\partial p_b} = \phi_i \boldsymbol{a_{ia}}'(A_i' M_i A_i)^+ \boldsymbol{a_{ib}} \lesseqgtr 0, \ \forall a, b \in \mathcal{K} \tag{1.11}$$

The sparsity of the A_i and M_i matrices minimises individual product demand's reliance on the price vector of the entire product assortment (and vice-versa). Despite losing positive definiteness (and nice properties such as diagonal dominance) under the dimensionality constraint, the demand system is still well-behaved. To see this, note:

***Lemma 1.3**: $A_i(A_i' M_i A_i)^+ A_i'$ is symmetric and positive semi-definite.*

Symmetry and positive semi-definiteness in S_i means the (unconstrained) individual demand function is aggregate-monotonic, and therefore satisfies the Law of Demand (Amir et al., 2017; see also Hildenbrand, 1983).[6]

1.3.2 The Representative Consumer

In some settings, only market-level data is available, so researchers are compelled to work with aggregate demand. This raises a standard problem: aggregation may fail to discipline structural parameters, and aggregate parameters may not directly map to individual preferences in the consumer population. To resolve this, it is commonly assumed that (i) there exists a representative consumer whose utility-maximising demand for goods reproduces the observed aggregate demand system, and (ii) that the representative preference parameters can be tied to underlying consumer population preferences and are recoverable via estimation. Gorman (1953) provides conditions under which a representative consumer rationalisation is admissible for some degree of consumer heterogeneity.

However, those conditions abstain from how heterogeneity in consumer preferences implies different consumers have different reservation prices. And it is when individual consumers' consumption bundle choices face different binding non-negativity constraints

[6] It can also be shown that all own-price effects are strictly negative, i.e. $S_{i,aa} < 0$, $\forall a \in \mathcal{K}_i$, as M_i is positive definite and A_i has no zero-rows by construction.

that the representative consumer interpretation becomes especially delicate, even if each consumer's demand is linear within a given interior regime. Aggregation then generally combines demands from different regimes, so the resulting market demand need not be globally representable as the solution of a single representative consumer problem.

As such, a *representative consumer rationalisation* is generally not obtained in linear demand; no link can usually be established between a representative consumer's preferences and those of individual consumers.[7] This raises concerns for demand estimation, which aims to obtain structural preference parameters to analyse counterfactual scenarios. Without a rationalisation, counterfactual analysis using linear demand is inappropriate.

In this section, I consider if and when the model set up above admits a representative consumer rationalisation.

First, we need a common language for consideration sets across all consumers. Let $A_{K\times J}$, for $J > K$ be the aggregate consideration set, made up of the J unique columns across all A_i, such that, by the Minkowski sum, $\text{cone}(A) = \sum_{i=1}^{N} \text{cone}(A_i)$. Furthermore, let there exist a matrix $R_i \in \{0,1\}^{J\times J_i}$, $\forall i$, such that $R_{i,ab} = 1$ if the a-th column of A matches the b-th column of A_i, mapping out which baskets appear in the latter. Note then that $A_i = AR_i$, $\forall i$. Additionally, define $\boldsymbol{z}_i = R_i \hat{\boldsymbol{z}}_i$, with R_i playing the role of re-dimensioning $\hat{\boldsymbol{z}}_i$ from $J_i \times 1$ to $J \times 1$, with zeros in the new dimensions. The notation for the re-dimensioned choice bundle is $\boldsymbol{z}_i$. Then:

$$\boldsymbol{q}_i = T_i A_i \hat{\boldsymbol{z}}_i = T_i A R_i \hat{\boldsymbol{z}}_i = T_i A \boldsymbol{z}_i \tag{1.12}$$

This allows us to re-write the utility-maximisation problem of each consumer in terms of $\boldsymbol{z}_i$. The re-dimensioning cannot imply consumers make transactions they do not

[7] Choné and Linnemer (2020) argues this point explicitly in its discussion of heterogeneity in the quasilinear quadratic utility model.

consider, meaning it effectively imposes a new zero-purchase constraint on the optimisation: $(I_J - R_i R_i')\boldsymbol{z}_i = 0,\ \forall i$. Lastly, a dimensionality adjustment is applied to $\tilde{\boldsymbol{\lambda}}$ such that $\hat{\boldsymbol{\lambda}}_i' \hat{\boldsymbol{z}}_i = \boldsymbol{\lambda}_i' \boldsymbol{z}_i,\ \forall i$. Thus, in the same way we refer to $\boldsymbol{\lambda}_i$ as the non-negativity Lagrangean multiplier, we may refer to $\boldsymbol{\gamma}_i$ as the *zero-purchase* Lagrangean multiplier of consumer i's utility-maximisation problem. We will refer to the representative consumer's multipliers as $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ respectively.

We are now ready for the following proposition:

Proposition 1.3: *Let the $NN - LF - NORM$ bind. Assume there exists a symmetric positive definite matrix M such that each consumer's latent unconstrained Slutsky matrix is the M-restricted response on the consumer's feasible directions*:

$$S_i \boldsymbol{y} = \phi_i M^{-1} \boldsymbol{y}, \quad \forall \boldsymbol{y} \in \operatorname{span}(A_i),\ \forall i = 1, \ldots, N \tag{1.13}$$

Then, the aggregate demand:

$$\boldsymbol{Q}(\boldsymbol{p}) = \sum_{i=1}^{N} \boldsymbol{q}_i(\boldsymbol{p}) = A(A'MA)^{+}[A'(\sum_i \boldsymbol{\delta}_i + \phi \boldsymbol{p}) + \boldsymbol{\Lambda} + \boldsymbol{\Gamma}] \tag{1.14}$$

can be represented as the solution of a representative consumer's constrained maximisation problem in the same local functional class.

In other words, a binding consideration set and an agreed-upon shape for demand slopes across the individual consumers which consider them - i.e. what goods are substitutes/complements and to what extent - allow aggregation of preference heterogeneity in consideration sets, marginal utility of income, and initial marginal product utility as if reflecting the preferences of a single representative consumer.

The proposition assumes that consumers share a demand slope object M in the span(A_i), across all i in their latent *unconstrained* Slutsky matrix. "Unconstrained" is in reference to a setting where we hold fixed the inequalities that create kinks and corners, and focus only on the response in the linear feasible directions. This notion is helpful to isolate local effects away from wider considerations on the active set of consumers, which we will get to later. Heterogeneity in the individual unconstrained Slutsky matrices can still arise: either outside the set of goods defined in a given A_i (consumers may have different preferences over goods they do not buy); the considerations sets themselves; or the scalar factor ϕ_i. All remaining heterogeneity is found in the constraints, and therefore only appears for the constrained Slutsky matrix of each consumer.

This is a far stronger claim than previously made in the literature, and relies directly on the existence of consideration sets. All consumers, despite holding consideration sets with different dimensions, can have their optimisation problem be written using the same basket space. After stacking their first-order conditions to obtain an aggregate condition, the summed Kuhn–Tucker multipliers are shown to not automatically satisfy complementary slackness for a representative consumer. My contribution is to show that, because basket representations are not unique, you can reassign those multiplier terms into new aggregate multipliers that generate the same aggregate demand while restoring the representative consumer's KKT conditions. Thus, a combination of a basket choice bundle and a multiplier term can always be found that satisfies the KKT conditions for any given aggregate demand, recovering rationalisability.

Uniquely, this proposition directly addresses three of Eaton and Lipsey (1989)'s "awkward facts" about differentiated products, which were then argued to not be satisfied by representative consumer models: (i) that consumers purchase only a small subset of the products that are available in a given market; (ii) that consumers often have an approximate agreement as to product differentiation and what products are or not close substitutes/complements; and (iii) that consumers nonetheless decide upon different consumption bundles even where there are no relevant differences in income. **Proposition 1.3** claims a representative consumer rationalisation under precisely these conditions.

Proof: An aggregate consideration set A allows us to re-write the Lagrangian of each consumer i in terms of $\boldsymbol{z}_i$:

$$\begin{aligned}\mathcal{L}(\boldsymbol{z}_i) =& T_i\boldsymbol{z}_i'A'\boldsymbol{\delta}_i - \frac{1}{2}T_i^2\boldsymbol{z}_i'(A'MA)\boldsymbol{z}_i - \phi_i\Big(Y_i - T_i\boldsymbol{p}'A\boldsymbol{z}_i\Big) \\ &+ \boldsymbol{\lambda}_i'\boldsymbol{z}_i + \underbrace{\gamma_i'}_{\text{zero-purchase multiplier}}\boldsymbol{z}_i\end{aligned} \tag{1.15}$$

Note that the redimensioning and inclusion of a zero-purchase multiplier is addressed in the proof of **Lemma 1.2**, so $\eta_i = 0$ here too. The individual KKT conditions for $\boldsymbol{z}_i$ will then look something like this:

$$A'\boldsymbol{\delta}_i - T_i(A'MA)\boldsymbol{z}_i + \phi_iA'\boldsymbol{p} + \frac{1}{T_i}(\boldsymbol{\lambda}_i + \boldsymbol{\gamma}_i) = 0 \quad \boldsymbol{\lambda}_i \geq 0 \quad \boldsymbol{z}_i \geq 0 \qquad \boldsymbol{\lambda} \circ \boldsymbol{z}_i = \boldsymbol{0} \tag{1.16}$$

Stacking and collapsing the system of individual stationarity conditions, we get an aggregate:

$$A'(\sum_{i=1}^{N}\boldsymbol{\delta}_i) - (A'MA)(\sum_{i}T_i\boldsymbol{z}_i) + A'(\sum_{i=1}^{N}\phi_i)\boldsymbol{p} + \sum_{i=1}^{N}\frac{1}{T_i}(\boldsymbol{\lambda}_i + \boldsymbol{\gamma}_i) = 0 \tag{1.17}$$

However, this aggregate cannot be rationalised as the stationarity condition for a utility-maximising representative consumer, because the complementary slackness condition is not satisfied in general:

$$(\sum_{i=1}^{N}\boldsymbol{z}_i) \circ (\sum_{i=1}^{N}\boldsymbol{\lambda}_i) \neq 0 \tag{1.18}$$

This is because, in general, we have both $\sum_{i=1}^{N}\boldsymbol{z}_i > 0$ and $\sum_{i=1}^{N}\boldsymbol{\lambda}_i > 0$ in the population: at least one shopping basket is bought *and* not bought at least once. Yet for an identical stationarity condition produced by a representative consumer in $\boldsymbol{z}$ and $\boldsymbol{\Lambda}$, **Proposition 1.2** says the KKT conditions are satisfied. This is the general result known in the academic literature.

Our setup is slightly different however, incorporating consideration sets. Could we use these to find a way to conciliate this contradiction? Let us go step by step.

First, because of our condition that $\boldsymbol{q}_i = T_i A_i \boldsymbol{z}_i, \ \forall i$ (and therefore $\boldsymbol{Q} = TA\boldsymbol{z}$, for $\boldsymbol{z} = \frac{1}{T}\sum_{i=1}^{N} T_i \boldsymbol{z}_i$), there are infinitely many $\boldsymbol{z}$ one can have where this condition is satisfied: e.g. any $\boldsymbol{y}$ such that $\boldsymbol{Q} = TA(\boldsymbol{z}+\boldsymbol{y}), \ A\boldsymbol{y} = 0,$ and $\mathbf{1}'(\boldsymbol{z}+\boldsymbol{y}) = 1$ would do.

Second, note that, beyond the above condition on $\boldsymbol{z}$, for the purposes of understanding how shopping basket choice affects aggregate goods consumption $\boldsymbol{Q}$, it is not the stationarity condition in (1.17) that binds, but rather that which we obtain after we multiply both sides by $A(A'MA)^{+}$ to obtain the aggregate demand function:

$$\boldsymbol{Q} = A(A'MA)^{+}[A'(\sum_i \boldsymbol{\delta}_i + \phi_i \boldsymbol{p}) + \sum_i \frac{1}{T_i}(\boldsymbol{\lambda}_i + \boldsymbol{\gamma}_i)] \tag{1.19}$$

The $\boldsymbol{\Lambda}$ need not be in $\mathrm{col}(A')$. The same applies to a representative consumer's individual demand function, which we can write as:

$$\boldsymbol{q}^{rep} = A(A'MA)^{+}[A'(\sum_i \boldsymbol{\delta}_i + \phi \boldsymbol{p}) + \boldsymbol{\Lambda} + \boldsymbol{\Gamma}] \tag{1.20}$$

Our goal is therefore to determine, for $\phi = \sum_{i=1}^{N} \phi_i$, whether there are any vectors $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ for which

$$A(A'MA)^{+}[\sum_i \frac{1}{T_i}(\boldsymbol{\lambda}_i + \boldsymbol{\gamma}_i)] = A(A'MA)^{+}[\boldsymbol{\Lambda} + \boldsymbol{\Gamma}] \tag{1.21}$$

holds. Turns out, yes. To see this, note that the $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ are not separately identified. For infinitely many $\boldsymbol{z}$, say $\boldsymbol{z}^{\#}$, we can then set $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}^{\#}$ and $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^{\#}$ as follows:

$$\Lambda_j^{\#} = \begin{cases} 0 & \text{if } z_j^{\#} > 0 \\ \sum_i \frac{1}{T_i}(\lambda_{ij} + \gamma_{ij}) & \text{if } z_j^{\#} = 0 \text{ and } \sum_i \frac{1}{T_i}(\lambda_{ij} + \gamma_{ij}) \geq 0 \\ 0 & \text{if } z_j^{\#} = 0 \text{ and } \sum_i \frac{1}{T_i}(\lambda_{ij} + \gamma_{ij}) < 0 \end{cases}, \; \forall j = 1, \ldots, J \tag{1.22}$$

$$\mathbf{\Gamma}^{\#} = \sum_i \frac{1}{T_i}(\boldsymbol{\lambda}_i + \boldsymbol{\gamma}_i) - \mathbf{\Lambda}^{\#} \tag{1.23}$$

These generate the same $\boldsymbol{Q}$ as $\sum_{i=1}^N \frac{1}{T_i}\boldsymbol{\lambda}_i$ and $\sum_{i=1}^N \frac{1}{T_i}\boldsymbol{\gamma}_i$. It is trivial to see that $\mathbf{\Lambda}^{\#}$ obtained in such a way satisfy $\mathbf{\Lambda}^{\#} \circ \boldsymbol{z}^{\#} = 0$ and $\mathbf{\Lambda}^{\#} \geq 0$. Crucially, $\mathbf{\Gamma}^{\#}$ is what lets you keep the multiplier sum fixed while reallocating the nonnegative part into $\mathbf{\Lambda}^{\#}$ so that complementary slackness holds for the new values. This result applies locally - price changes imply a new set of multipliers to correct for. □

1.3.3 Discussion

While this is sufficient to confirm the existence of a representative consumer rationalisation, it is not enough to allow for separable identification of preference parameters - for a general A, $\boldsymbol{\delta} = \sum_{i=1}^N \boldsymbol{\delta}_i$ is not uniquely identified. However, note that:

Lemma 1.4: *If A is full row rank, then $A(A'MA)^{+}A' = M^{-1}$.*

Then:

$$\begin{aligned} \boldsymbol{Q} =& A(A'MA)^{+}[A'(\sum_i \boldsymbol{\delta}_i + \phi_i \boldsymbol{p}) + \sum_i \frac{1}{T_i}(\boldsymbol{\lambda}_i + \boldsymbol{\gamma}_i)] \\ =& M^{-1}(\boldsymbol{\delta} + \phi \boldsymbol{p}) + A(A'MA)^{+}[\mathbf{\Lambda} + \overline{\boldsymbol{\gamma}}] \end{aligned} \tag{1.24}$$

for $\boldsymbol{\delta} = \sum_i \boldsymbol{\delta}_i$ and $\phi = \sum_{i=1}^N \phi_i$.

In fact, it is for this reason that we have that $J > K + 1$ in the proof earlier. If A is full row rank, null$(A) = J - K$; because $\mathbf{1}'\boldsymbol{z} = 1$, the total number of dimensions of the linear space of $\boldsymbol{z}$ outside $cone(A)$ is reduced by 1. We return to this full rank assumption later.

We now have relevant conditions under which we can learn of underlying population preference parameters using only market-level information. We make use of this in Chapter 2. Locally, where income effects are not relevant, proportional demand slopes under rational inattention are sufficient for a population of utility-maximising individual consumers to have their aggregate demand rationalised as that of a single utility-maximising representative consumer.

1.3.4 Equilibrium Analysis

Having built-up an understanding of how linear aggregate demand operates, we consider now its implications for the outcomes of (Bertrand) price competition. We will account for two extreme market structures: that (i) of a finite number of single-product firms, and (ii) of a single multi-product monopoly. Both settings will be shown to deliver well-behaved demand with a unique equilibrium.

Assume first that all K differentiated products in the assortment are sold by K different single-product firms. For simplicity of exposition, marginal costs are set to $c_i = 0, \forall i \in \mathcal{K}$. The representative consumer takes prices as given. All results are shown assuming interior solutions only for simplicity.

Proposition 1.4: *When each of the K differentiated products is sold by a separate* *firm engaging in Bertrand (Nash) competition, the unique equilibrium price vector is*:

$$\boldsymbol{p}^* = -\frac{1}{\phi}[\Omega + A(A'MA)^+A']^{-1}A(A'MA)^+A'\boldsymbol{\delta} \tag{1.25}$$

for $\Omega = \text{diag}(\boldsymbol{a_1}'(A'MA)^+\boldsymbol{a_1},\ \ldots,\ \boldsymbol{a_K}'(A'MA)^+\boldsymbol{a_K})$.

Proof: Consider a given firm i's objective function:

$$\pi_i(\boldsymbol{p}) = p_i q_i(\boldsymbol{p}) = p_i\Big[a_i'(A'MA)^+A'(\boldsymbol{\delta} + \phi\boldsymbol{p})\Big] \tag{1.26}$$

The set of first-order conditions for this problem across all firms may therefore be defined as

$$0 = A(A'MA)^+A'\boldsymbol{\delta} + 2\phi\Omega\boldsymbol{p} + \phi(A(A'MA)^+A' - \Omega)\hat{\boldsymbol{p}} \tag{1.27}$$

To find the Bertrand (Nash) equilibrium, we look for the fixed point $\boldsymbol{p}^*$ such that $\boldsymbol{p} = \hat{\boldsymbol{p}} = \boldsymbol{p}^*$ - i.e. that point which defines the best response as a function of all competing firms' strategies. Uniqueness is enabled by the following property:

Lemma 1.5*:* $\Omega + A(A'MA)^+A'$ *is invertible.* □

It then follows that:

Proposition 1.5*: In the single-product firms setting, the unique optimal demand* $\boldsymbol{Q}(\boldsymbol{p}^*)$*, is defined by the expression:*

$$\boldsymbol{Q}(\boldsymbol{p}^*) = \Omega(\Omega + A(A'MA)^+A')^{-1}A(A'MA)^+A'\boldsymbol{\delta} \tag{1.28}$$

Proof: Result follows from substituting $\boldsymbol{p}^*$ onto $\boldsymbol{Q}(\boldsymbol{p}) = A(A'MA)^+(\boldsymbol{\delta} + \phi\boldsymbol{p})$.

What if all K differentiated products in the assortment are sold by a single multi-product monopolist retailer?

Proposition 1.6: *When each of the K differentiated products is sold by a single, multi-product, price-setting monopolist, a multiplicity of price equilibria arises:*

$$\boldsymbol{p}^* = -\frac{1}{2\phi}\boldsymbol{\delta} + (I - W^+W)\boldsymbol{y} \tag{1.29}$$

with

$$W = \Omega + A(A'MA)^+A' + A(A'MA)^+A' \circ G = 2A(A'MA)^+A' \tag{1.30}$$

for $\boldsymbol{y} \in \mathbb{R}^K$ and G is a symmetric hollow $K \times K$ ownership matrix where each element $G_{ab} = 1$ if $a, b \in S$ (i.e. owned by the same firm), 0 otherwise. In this case, $G_{monopolist} = \boldsymbol{1}\boldsymbol{1}' - I$.

Proof: A multi-product firm determines a vector of prices for its goods by maximising the following profit function:

$$\pi_S(\boldsymbol{p}) = \sum_{i \in S} p_i[a_i'(A'MA)^+A'(\boldsymbol{\delta} + \phi\boldsymbol{p})] \tag{1.31}$$

for S the subset of products $i = 1, \ldots, K$ sold by said firm. The FOC can be expressed in matrix form as:

$$0 = A(A'MA)^+A'\boldsymbol{\delta} + \phi(2\Omega + 2(G \circ A(A'MA)^+A'))\boldsymbol{p} + \phi((A(A'MA)^+A' - \Omega) - G \circ A(A'MA)^+A')\hat{\boldsymbol{p}} \tag{1.32}$$

To find the Bertrand (Nash) equilibrium, we operate in a manner identical to that pursued to find the equilibrium price in the single-product firms case. The $(I - W^+W)\boldsymbol{y}$ which arises as a result of the pseudo-inverse implies a multiplicity of mappings from initial marginal product utility to equilibrium prices. □

The inclusion of an ownership matrix is problematic for the identification of a unique price equilibrium in more general market structures, as we cannot say anything *ex-ante* about the definiteness of G. In fact, without additional assumptions, no closed-form solutions for multi-product oligopoly case exist for this model (more on this in the Appendix). However, for the monopoly case in particular, the multiplicity of price equilibria are not just payoff-equivalent, but *equilibrium-quantity equivalent*:

Proposition 1.7: *In the monopoly setting, the unique optimal demand $\boldsymbol{Q}(\boldsymbol{p}^*)$, is defined by the expression:*

$$\boldsymbol{Q}(\boldsymbol{p}^*) = \frac{1}{2}A(A'MA)^{+}A'\boldsymbol{\delta} \tag{1.33}$$

while the aggregate profit and consumer surplus take on the following unique closed-form expressions:

$$\Pi = \sum_{i}^{N} \pi_i = \boldsymbol{Q}'(\boldsymbol{p}^*)\boldsymbol{p}^* = -\frac{1}{4\phi}\boldsymbol{\delta}'A(A'MA)^{+}A'\boldsymbol{\delta} \tag{1.34}$$

$$CS = \boldsymbol{Q}'(\boldsymbol{p}^*)\boldsymbol{\delta} - \frac{1}{2}\boldsymbol{Q}'(\boldsymbol{p}^*)M\boldsymbol{Q}(\boldsymbol{p}^*) + \phi\Pi = \frac{1}{8}\boldsymbol{\delta}'A(A'MA)^{+}A'\boldsymbol{\delta} \tag{1.35}$$

Proof: The proof is trivial once we substitute $\boldsymbol{p}^*$ onto $\boldsymbol{Q}(\boldsymbol{p}) = A(A'MA)^{+}A'(\boldsymbol{\delta}+\phi\boldsymbol{p})$ as well as onto the representative consumer's utility function and the monopolist's aggregate profit function. □

1.4 Comparative Statics

In what follows, for simplicity, we work with the NN-LF condition:

$$\boldsymbol{Q} = A\boldsymbol{z} \qquad \& \qquad \boldsymbol{z} \geq 0 \tag{NN-LF}$$

We leave additional adjustments to account for the zero-purchase constraint imposed by $\overline{\gamma}$ ($(I - R_iR_i')\boldsymbol{z}_i = 0,\ \forall i$) in the next Chapter. Assume nonetheless **Proposition 1.3** holds.

Let us consider how the NN-LF condition impacts aggregate demand. By **Proposition 1.1**, we can separately address vector space restrictions from the the broader condition which includes choice vector non-negativity. We do so by varying the row rank of A and relying on **Lemma 1.4**. When the full NN-LF condition binds,

$$\begin{aligned} \boldsymbol{Q} &= A(A'MA)^{+}A'(\boldsymbol{\delta} + \phi \boldsymbol{p}) \;\; s.t. \;\; \boldsymbol{z} \geq 0 \\ &= A(A'MA)^{+}A'(\boldsymbol{\delta} + \phi \boldsymbol{p}) + A(A'MA)^{+}\boldsymbol{\Lambda}(\boldsymbol{\delta}, \boldsymbol{p}) \end{aligned} \tag{1.36}$$

When only the NN portion binds, $span(A) = \mathbb{R}^K_+$, $A(A'MA)^+A' = M^{-1}$, and

$$\begin{aligned} \boldsymbol{Q} &= M^{-1}(\boldsymbol{\delta} + \phi \boldsymbol{p}) \;\; s.t. \;\; \boldsymbol{z} \geq 0 \\ &= M^{-1}(\boldsymbol{\delta} + \phi \boldsymbol{p}) + A(A'MA)^{+}\boldsymbol{\Lambda}(\boldsymbol{\delta}, \boldsymbol{p}) \end{aligned} \tag{1.37}$$

The non-negativity multiplier $\boldsymbol{\Lambda}$, has positive elements when the corresponding basket coordinate binds at the lower bound, e.g. $\lambda_i > 0$ if $z_i = 0$. The term $A(A'MA)^+$ then maps it into product space. We may call the full expression - $A(A'MA)^+\boldsymbol{\Lambda}$ - a corner-solution adjustment. When $A(A'MA)^+\boldsymbol{\Lambda}(\boldsymbol{\delta}, \boldsymbol{p}) = \boldsymbol{0}$, closed-form interior solutions hold. Significantly, $A(A'MA)^+\boldsymbol{\Lambda}(\boldsymbol{\delta}, \boldsymbol{p})$ is a function of price and the initial marginal product utility; we return to this later to discuss its implications for how to address price endogeneity in an empirical setting.

The key difference between the equations above is that the unconstrained Slutsky matrix either depends on $A(A'MA)^+A'$ or M^{-1}. Can we say anything about how they differ structurally? Yes:

Proposition 1.8: *Where the LF condition binds, relative to M^{-1}, $A(A'MA)^+A'$ has (weakly) lower-valued diagonal elements.*

Proof: The trick for this proof is to use congruence. Note that $M^{-1} = M^{-1/2}IM^{-1/2}$ and $A(A'MA)^+A' = M^{-1/2}PM^{-1/2}$, for $P = \tilde{A}(\tilde{A}'\tilde{A})^+\tilde{A}'$ where $\tilde{A} = M^{1/2}A$ and P is an orthogonal projection onto the column space of $\tilde{A}$. By the aforementioned properties of the M matrix, this means M^{-1} and I are congruent, and so are $A(A'MA)^{+}A'$ and P. For our purposes, this allows us to compare P and I knowing that said comparison extends to their congruents. For example, consider the diagonal elements of P. The definiteness properties of I and P's congruents extend to them: $\boldsymbol{x}'P\boldsymbol{x} \leq \boldsymbol{x}'I\boldsymbol{x}, \forall \boldsymbol{x}$. Since $\boldsymbol{x}'I\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{x} = ||\boldsymbol{x}||^2$, $P_{ii} = \boldsymbol{e}_i'P\boldsymbol{e}_i \leq \boldsymbol{e}_i'\boldsymbol{e}_i = ||\boldsymbol{e}_i||^2 = 1$. As all diagonal elements of I are equal to 1, this means all diagonal elements of P are weakly smaller than I's, and all diagonal elements of $A(A'MA)^{+}A'$ are weakly smaller than M^{-1}'s (one need only think of $\boldsymbol{x} = M^{-1/2}\boldsymbol{e}_i, \forall i$). Each individual diagonal element of $A(A'MA)^{+}A'$ is weakly smaller than the diagonal elements of M^{-1}, so the average of these is also smaller. □

This is an *attenuation effect.* Where consumers consider and choose over combinations of goods, individual goods and their characteristics drive their own demand (weakly) less, as that will depend on the wider set being purchased. In these cases, own-price makes (weakly) less of an impact than it would otherwise. This result can be extended also to the non-negativity condition, which imposes an attenuation effect of its own.

Corollary 1: *Where the non-negativity condition binds, own-price effects are (weakly) smaller in magnitude.*

Proof: The NN condition introduces an active set of baskets with $z_j > 0$ and separates it from an inactive set where $z_j = 0$. This effectively replaces the column space of A with a smaller (active) subspace in the column space of $A_{z_j>0} \subseteq A$. As we saw, orthogonal projection onto a smaller subspace weakens the length of any component's projection, thus leading to diagonal entries (weakly) smaller in magnitude. □

Important questions remain. No stronger claim can be made regarding the exact change in the characteristics or magnitude of the off-diagonal elements of $\phi A(A'MA)^{+}A'$ relative to ϕM^{-1}. Similarly, not much can be said of the impact of the different characteristics of the model on its equilibrium outcomes. To address these limitations, I run a Monte Carlo simulation. This allows us to systematically generate random environments, solve for equilibrium quantities and prices under different market structures, and evaluate different consideration set structures. While not constituting solid proof, these outcomes provide some degree of confidence as to the direction of some otherwise hard-to-prove effects, and constitute hypotheses for further research. The simulation is run over 1000 draws, and is divided into three steps: primitive sampling, Hessian matrix formation, and equilibrium computations.

In the first step, the representative consumer's consideration set A is produced from a baseline of 10 goods and 51 baskets (roughly 5% of all feasible combinations). After correcting for zero- and non-unique columns and rows via non-negative linear combinations of the remaining, up to 10 columns and rows are also added in such a way. The average draw has 14.8 goods and 55.5 baskets, for an average rank of 10 and an average 14.3 missing standard basis vectors. The marginal-utility-of-income coefficient ϕ is set to -0.5.

In the second step, the consideration set A is leveraged to produce the Hessian matrix M. To do so, I implement the approach in Tian et al. (2021) to produce said matrix based on co-purchase regularities (approach discussed in greater detail in the next Chapter). The approach allows both A and M to be indirectly inferred from the same underlying simulated preferences. The inverse matrix itself is 50% said specification, 50% randomised, to reflect possible measurement error.

In the last step, I compute four different models, underpinned by two market structures: multi-good monopoly and single-good monopolistic competition. The models are: baseline, with outcomes derived from an unconstrained Slutsky matrix where the NN-LF condition does not bind; unconstrained, where only the vector space restriction does; and lastly two setups with outcomes derived from the constrained Slutsky matrix - one where only non-negativity binds, and one where the NN-LF condition binds in full.

After each simulation draw, all models are compared across a range of characteristics. Given the price elasticity of demand in each draw, I determine own-price elasticity, cross-price elasticity, average magnitude of substitutes' and complements' cross-price elasticities, and the average number of complementary and substitute products per good, well as outcomes: average demand, consumer surplus, and aggregate profits.

Table 1.1: Simulation of consumer choice counterfactuals

Metric	**Unconstrained**	**Baseline**	**Relation**	**% draws**
Avg own-price effect	−0.37	−0.53	< (abs.)	100
Avg cross-price effect	0.008	0.0186	<	100
Avg substitution effect	0.0339	0.0323	>	75.9
Avg complementary effect	−0.0502	−0.0199	> (abs.)	100
Avg number of substitutes	4.7	5.2	<	88.3
Avg number of complements	2.1	1.7	>	88.3
Metric	**Const. (NN-LF)**	**Const. (NN)**	**Relation**	**% draws**
Avg own-price effect	−0.37	−0.53	< (abs.)	100
Avg cross-price effect	0.008	0.0186	<	100
Avg substitution effect	0.0339	0.0323	>	75.6
Avg complementary effect	−0.0502	−0.0199	> (abs.)	99.9
Avg number of substitutes	4.7	5.2	<	88
Avg number of complements	2.1	1.7	>	88.2
Firm consumer surplus	89.6	178	<	100
Monopoly consumer surplus	106.5	107.3	<	94.8
Firm aggregate profit	371.4	391	<	56.6
Monopoly aggregate profit	425.9	429.3	<	100
Firm aggregate demand	2.60	3.73	<	100
Monopoly aggregate demand	2.88	2.90	<	93.7

Notes: "Relation" compares the outcomes of consumer choice under consideration set constraints to the unconstrained setting value. Rows marked "(abs.)" compare absolute values. "% draws" is the percentage of simulation draws in which the stated relation (weakly) holds. No metric has an equality relation between models over 15% of all draws.$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

A few results stand out. Firstly, the constraints imposed on aggregate demand by a representative consumer's corner solutions do not appear to qualitatively affect outcomes.[8] Instead, a set of suggestive relationships between models bounded or not by the vector space (or LF) restriction emerges.

Unsurprisingly, the absolute value of the average own-price effect is (weakly) lower when choice is constrained by the representative consumer's consideration set restriction on the consumption vector space. This is as predicted in **Proposition 1.8**, though slightly stricter.

More novel: the average cross-price effect is (weakly) lower when choice is constrained by the representative consumer's consideration set restriction on the consumption vector space. This constitutes one of the more elusive proofs of interest to us. Overall, the attenuation effect found in own-price effects appears to extends to cross-price effects.

The following offers some suggestions on the drivers of this result: the average magnitude of cross-price effects from complementary goods is (weakly) greater when choice is constrained by the representative consumer's consideration set restriction on the consumption vector space. Restricting consumption bundles to a consideration set drives an increase in complementarity, caused by regular co-consideration between goods. These appears to be another side of the attenuation effect - product demand is driven by demand of those goods it is co-considered with most often.

This is paired with a few rarer (but not so rare) occurrences. The average magnitude of cross-price effects from substitute goods is tendentially greater when choice is constrained by the representative consumer's consideration set restriction on the consumption vector space. On one hand, substitution effects grow stronger, much like complementarity. This

[8]This may nonetheless differ depending on the number and type of consumers which induce such corners.

is intuitive. Conditional on shared complements, the new competitive margin is in what to pair said complement with. In LF-unconstrained models, the decision to buy one good has a broader impact on those not bought; but in a setting where pairing one good with another more directly implies a decision not to buy yet another good that the first is often co-considered with strengthens the rivalry between the chosen and not-chosen good. On the other hand: the average number of complementary goods per product is generally greater when choice is constrained by the representative consumer's consideration set restriction on the consumption vector space.; and the average number of substitute goods per product is generally lower when choice is constrained by the representative consumer's consideration set restriction on the consumption vector space. In other words, on the net, the complementarity effect takes precedence in the cross-price outcome.

Amongst the market outcomes, note the following. Consumer surplus is (weakly) lower when choice is constrained by the representative consumer's consideration set restriction on the consumption vector space, in the single-product firm setting; but only tendencially lower in the same circumstances in the multi-product monopoly setting. Note also that product ownership concentration appears to lead to greater consumer surplus where the consideration set restriction on the consumption vector space binds in a majority of simulation draws, a result virtually absent in product choice. This suggests that the above-mentioned complementarities created by co-consideration are best taken advantage of in concentrated markets where incentives are internalised.

Aggregate profit is tendencially lower when choice is constrained by the representative consumer's consideration set restriction on the consumption vector space, in the multi-product monopoly setting. Surprisingly, however, under the same conditions, single-product firms yield greater aggregate profits in only a small majority of cases. This may be driven by the two possibilities discussed above: weaker own-price effects result in higher pricing power and thus higher prices; smaller cross-price effects - which as we saw are driven by a greater (lesser) number of complements (substitutes) and complementary (substitution) effects - diminish competitive concerns. The strategic complementary characteristic of Bertrand competition does the rest. Nonetheless, aggregate profits are always greater under multi-product monopoly than in single-product oligopolistic settings.

Lastly, average demand is tendencially lower when choice is constrained by the representative consumer's consideration set restriction on the consumption vector space, across both single-product oligopolistic and multi-product monopoly settings. Monopoly aggregate demand is greater in the multi-product monopoly, in keeping with our discussion above.

The rational inattention literature consistently finds that behavioural restrictions to consumer choice reduce price elasticity (Matějka and MacKay, 2015; Caplin, Dean, and Leahy, 2019). This is intuitive for our setting too. Where choices are restricted by the NN-LF, for every consideration set alternative its next best alternative is on average worse than if there was no such restriction on the consumption bundles. Unlike in the existing literature, this effect is not necessarily driven by an "information gap", but can instead be a consequence of the underlying structure of preferences itself.

On the whole, when the vector space restriction of the consideration set binds, own-price effects are weaker than they would otherwise be, as goods' demand becomes relatively more dependent on those goods with which they are often co-purchased. Often co-considered goods behave as complements, while goods often paired with the same products face substitution effects conditional on these. From these, consumer choice under more realistic consideration set constraints also translates into worse market outcomes for both consumers and firms than previous theory otherwise assumed; and it favours product ownership concentration at a greater clip, with the average draw repeatedly finding greater aggregate demand, consumer surplus, and aggregate profits in this setting.

1.5 Illustrative Example

To wrap up this Chapter, we conclude with an illustrative example of what we have learned. At a base level, some goods effectively become more or less attractive, closer or farther substitutes, or even complementary, as cross-price effects become intertwined with often-co-purchased goods through considerations set constraints. Let

$$A_{\substack{\{red\ onion,\ avocado,\\ lime\}\times choices}} = \begin{pmatrix} 2 & 0 & 1 & 2 \\ 2 & 2 & 2 & 4 \\ 0 & 2 & 1 & 2 \end{pmatrix} \tag{1.38}$$

One approach for producing a substitute matrix consistent with the observed consideration set $A_{\{red\ onion,\ avocado,\ lime\}\times choices}$ (Tian et al., 2021) produces the following matrices:

$$M^{-1} = \begin{pmatrix} 1.140 & -0.107 & -0.370 \\ -0.107 & 1.030 & -0.107 \\ -0.370 & -0.107 & 1.140 \end{pmatrix},\ A(A'MA)^{+}A' = \begin{pmatrix} 0.884 & 0.257 & -0.627 \\ 0.257 & 0.514 & 0.257 \\ -0.627 & 0.257 & 0.884 \end{pmatrix} \tag{1.39}$$

Notice how, despite all three items behaving as substitutes in the first matrix, avocado becomes a complement to all in the second matrix. Red onions and lime, meanwhile, become stronger substitutes. This seems to be driven by avocado being in all consideration set alternatives. On what guacamole is concerned, avocado is more likely to be the one driving the purchases - sometimes with red onion, sometimes with lime, sometimes with both. All feasible ways for the consumer to adjust what they buy run through it. Once choice is constrained, this fact makes avocado a complement to both red onions and lime; conditional on it, the relevant margin is what to pair it with, milk or pasta. This residual rivalry explains the strengthening of substitute effects. The flipping of avocado from a substitute to a complement explains the rest. Effectively, regular co-consideration appears to drive complementarities, while strengthening substitution between goods which share co-considered products.

Following the example, we can also calculate aggregate profits, consumer surplus, and average price and demand across goods and market structures. At the Bertrand (Nash) equilibrium across both single-product firms and monopolist pricing:

Table 1.2: Welfare outcomes under various market structure and consideration set constraint specifications.

Π/CS	**NN-LF**	**Baseline**
K firms	20.22/0.78	20.36/1.57
Monopolist	20.54/1.03	21.41/1.07

Table 1.3: Average market outcomes under various market structure and consideration set constraint specifications.

$\overline{\boldsymbol{p}^*}/\overline{\boldsymbol{q}^*}$	**NN-LF**	**Baseline**
K firms	11.269/0.598	7.865/0.863
Monopolist	9.997/0.685	10/0.714

All of these match what we observed in the Monte Carlo simulation.

1.6 Discussion and Conclusion

This Chapter has developed a novel way of thinking about consumer choice in differentiated-products markets where purchases commonly involve multiple goods, multiple units, and combinations across categories. The central idea is that consumers do not simply choose quantities of individual goods from the full product space, but instead choose over finite consideration sets composed of shopping baskets. Modelling choice in this way makes it possible to accommodate realistic limits on attention within a tractable framework.

The main theoretical finding is that, even when basket choice itself is not uniquely pinned down, the induced demand over goods remains unique and linear. It can then be shown that consideration-set-constrained choice can provide a workable microfoundation for high-dimensional continuous demand systems. If consumers share proportional local demand slopes and the aggregate consideration set is sufficiently rich, then aggregate demand can be rationalised by the behaviour of a representative consumer within the same functional class, despite heterogeneity in preferences, price sensitivity, and consideration sets. This substantially relaxes the homogeneity requirements usually associated with representative-consumer results and helps address a long-standing concern in the linear-demand literature. It also provides the conceptual bridge to empirical work with market-level data, where the researcher observes aggregate outcomes but not the full structure of individual choice.

The Chapter also characterises how consideration sets shape equilibrium outcomes. When consumer choice is constrained by the structure of shopping baskets, demand becomes more dependent on patterns of joint consideration and joint purchase. Goods that are regularly co-considered may behave as complements, while goods that compete for association with the same baskets may become stronger substitutes conditional on those links. More generally, the vector-space restriction implied by consideration sets attenuates own-price responses and, in the simulations, tends also to weaken average cross-price effects. The Monte Carlo evidence and an illustrative example further suggest that constrained choice lowers average demand and consumer surplus, and often lowers profits as well, while also creating cases in which greater product ownership concentration may be welfare-improving. These results point to the importance of modelling demand at the level of combinations of goods rather than treating all interactions as arising solely from standalone products.

Incidentally, the consideration set approach in this Chapter happens to mirror Lancaster (1966) and its treatment of "activities", the name given in that paper to shopping baskets. This Chapter can therefore also be understood as contributing to shedding light on this less well-known feature of that seminal paper; making said analysis mathematically tractable; and applying it to further our understanding of consumer choice over large sets of multi-product cross-category alternatives.

Most of the findings that have been presented relate to the vector space restriction. The restriction will always bite if consumers consider more products across shopping baskets than the total number of shopping basket alternatives considered. Outside of this setting, how often is the aggregate consideration set less than full row rank in reality? Even if linear dependence is the case for a single individual's consideration set, aggregation into a representative consumer will likely change this. Even if individual purchase patterns are driven by latent factors that create linear dependency, aggregation may bring the aggregate consideration set to full rank. Each individual might even be influenced by a similar set of factors, but should the factors hold different weights depending on the consumer, this heterogeneity may blur the low-rank structure. An aggregated consideration set with contributions from many individuals with differing factor loadings will have increasing effective diversity (or rank). In other words, while each individual's consideration set might be low rank, averaging over many individuals could add enough independent variation that the aggregate no longer possesses said dependencies. We explore this possibility at length with real-life data in Chapter 2.

2

Empirical Estimation of Aggregated Linear Demand Models

Contents

2.1	**Introduction**	**36**
2.2	**Literature Review**	**40**
2.3	**Data**	**45**
2.4	**Empirical Application**	**50**
2.4.1	Empirical Specification	50
2.4.2	Identification Strategy	59
2.5	**Results**	**60**
2.6	**Discussion and Conclusion**	**72**

2.1 Introduction

Demand estimation provides a picture of how buyers react to prices and product features, allowing us to evaluate market power, assess mergers, and track spillovers.[1] It

[1] I have no relevant or material financial interests that relate to the research described in this Chapter. I gratefully acknowledge financial support by the Department of Economics at the University of Oxford

also lets us gauge the effects of taxes, subsidies, or new rules before they are put in place, allowing for counterfactual analysis of options and trade-offs.

In the standard model of choice behaviour in Economics, choice is characterised as the result of perception, driving preferences, converging into a process which produces a choice (McFadden, 2001). Perception and preferences are defined over goods and their attributes; goods are also the available choices. Textbook rational consumers then consider every feasible consumption bundle of these goods, selecting the combination which yields the greatest utility affordably (see e.g. Mas-Colell, Whinston, and Green, 1995).

This base setup raises two problems for demand estimation. First, the assumption that consumers consider every feasible consumption bundle is unrealistic. For example, guacamole is a popular spread which most commonly mixes avocado, red onion, and lime, items which are otherwise rarely bought in pairs; consumers tend to only consider either the combination or the individual goods, but not the pairs. This also challenges the idea of preferences and choices over goods. The choice to buy the combination of avocados, red onions, and limes is lower-dimensional than their separate choice - i.e. demand for each good is dependent on that of the others. This dependence may have implications for the outcomes of consumer choice.

Second, even when adequate, standard assumptions remain computationally infeasible. Gandhi and Nevo (2021) discuss at length the core challenges of demand modelling and estimation for differentiated goods. In practice, empirical applications target discrete, individual-good choice for a few goods within a product category; this approach can also

and the Fundação para a Ciência e Tecnologia of the Ministry of Education, Science and Innovation (Portugal), as well as data sharing by SONAE. The data agreement involves a request for review of the findings prior to their release. I benefited from thoughtful conference discussions by Leonardo Madio, Anne-Christine Barthel, and Alminas Zaldokas, as well as valuable comments from participants of the NIE PhD Symposium 2024, the 13th Oligo Workshop, the Research Jamboree at the University of Oxford, the Lisbon Meetings in Game Theory and Applications #14, EARIE 2025, the CEPR Paris Symposium 2025, and MaCCI 2026.

be extended to account for inattention, bounded rationality, search costs, and (unobserved) constraints (see e.g., Abaluck and Adams, 2021). Yet discrete-choice models are restrictive in their own way. The allowable set of consumption bundles is simultaneously too small and too large: cross-category complementarity is omitted (e.g. butter and bread, milk and cereal) and many goods are included despite being rarely considered without others (e.g. add-on sales, impulse purchases, or nonfood items at grocery stores). As a whole, this underestimates price elasticities, producing excessive estimates of market power and understating passthroughs.

In this Chapter, I proceed from the theoretical model introduced in Chapter 1, and propose an approach that addresses both the technical and computational challenges of estimating continuous demand systems. Incorporating Chapter 1's analysis of consideration sets onto demand estimation enables modelling under more realistic conditions. To estimate linear demand, I use point-of-sale data from across a sample of grocery stores owned by a supermarket chain in Portugal, for the period from February 2020 to February 2023. The data covers almost 30000 goods and more than 500 product categories. The representative consumer's consideration set is derived from the set of unique transactions observed during this period.

In Chapter 1, I discussed how, when binding, consideration sets affect linear demand in two ways through the NN-LF condition. For sufficient consideration set heterogeneity across consumers, I show how the interior Slutsky matrix - that is the matrix of price effects not accounting for Lagrangean constraints on $\boldsymbol{z}$ - behaves independently of the aggregate consideration sets. However, unlike in Chapter 1, these constraints play a key role by introducing a form of price endogeneity in the system. They create the well-known piecewise linearity in aggregate demand - different consumers are constrained differently, such that a different subset of these are active for any given set of prices. Ignorance of such constraints causes prices to be correlated with the error term because, e.g. as prices rise, the necessary adjustment to force negative demands back to zero also rises. The proposed empirical approach in this Chapter addresses this, allowing a separation between aggregate price effects due to price elasticity from those due to changes in the set of active consumers.

I successfully recover mark-up measurements across the product assortment. Between 2020 and 2023, I show mark-ups remain relatively stable across all major product categories. I document revenue-weighted mean mark-ups averaging $\approx 8.5\%$, with a peak in the third quarter of 2022 ($\approx 10.5\%$) - levels below those previously documented (e.g. Dopper, MacKay, Miller, and Stiebale, 2024). Benchmarking against U.S. grocery survey margins on a four-quarter moving average basis reveals near-identical timing and volatility.

Robust measures of mark-up distribution over the sample period show that changes are statistically insignificant except for an $\approx 7.4\%$ rise among top-percentile products, which seems to start to some extent two quarters prior, coincident with the CPI-price-index decoupling. Importantly, the 2022 increase is not driven by supermarket heterogeneity. To study the drivers of this late mark-up rise, I study transitions between mark-up percentiles, finding high persistence with mild mean reversion, but weakened persistence in mid-to-late 2022. This is consistent with shifting tastes and emergence of new high-mark-up goods. Analysis of fixed-basket mark-up indices further indicate reallocation by the third quarter of 2022.

The Chapter is structured as follows. Section 2.2 reviews the relevant literature. Section 2.3 introduces the data. Section 2.4 describes our application to study market power from intra-supermarket competition, including our empirical specification and identification strategy. Section 2.5 analyses the estimation results. Section 2.6 concludes.

2.2 Literature Review

In empirical microeconomics, more specifically industrial organisation, early attempts to estimate choice behaviour in continuous demand settings (e.g. Theil, 1965; Christensen, Jorgenson, and Lau, 1975; Deaton and Muellbauer, 1980) aimed to allow demand estimation across large numbers of goods flexibly, but faced some difficulties (Gandhi and Nevo, 2021). In these models, as the number of alternatives grows larger, the number of own- and cross-price effect parameters grows (proportional to) quadratically in these, producing a curse of dimensionality. Price endogeneity also becomes nearly impossible to address, due to collinearity and the need to find enough exogenous (and non-collinear) instruments to separately identify price effects. Other issues, such as addressing consumer heterogeneity and the entry of new goods, are also relevant difficulties.

Constant Elasticity of Substitution (CES) demand (Spence, 1976; Dixit and Stiglitz, 1977) and the logit model (introduced by McFadden, 1973) offered responses to these limitations, with significantly different utility functional forms and assumptions. The latter has become especially popular in empirical industrial organisation; Berry, Levinsohn, and Pakes (1995) remains the key reference for demand estimation via discrete-choice models. Other significant contributions in this tradition include Nevo (2001) for supermarket scanner data; and Hausman, Leonard, and Zona (1994) on instrumental variables for price endogeneity. Discrete-choice models enable modelling individual choice under consumer heterogeneity and allow for correlations in unobserved factors. However, their fit for applications over large choice sets is also limited by their own curse of dimensionality. This poor scalability is compounded by computational difficulties (see, e.g., Lanier, Large, and Quah, 2023) and concerns about large market asymptotics (Armstrong, 2016). Overcoming these difficulties remains a work in progress (see e.g. Iaria and Wang, 2019; Ershov, Laliberté, Marcoux, and Orr, 2024).

Researchers' ability to study competition and market structure in settings with large choice sets has thus been limited. Dopper, MacKay, Miller, and Stiebale (2024) use scanner data across over 100 consumer product categories to estimate demand and recover mark-up trends. The authors apply a random coefficient logit demand system and a covariance-restrictions approach to each individual category and estimate demand for each in isolation. Fosgerau, Monardo, and de Palma (2024) achieve further scalability and flexibility in a similar setting, employing a generalised inverse logit model, optimised for computational efficiency by leveraging market segmentation. Their approach effectively allows for a much greater number of alternatives than previous versions of logit model by adjusting the model such that it can be estimated via linear instrumental variables regression with market-level data. This linear approach echoes that of the present Chapter. However, it does not explicitly overcome the concerns regarding large market asymptotics which discrete choice raises, nor does it directly account for consumers considering combinations of goods over individual product purchases.

The present Chapter instead builds on the earlier attempts. In particular, we pursue an alternative framework based on consideration sets for thinking about multi-good cross-category purchases. This differs from the multi-level demand systems of Gorman (1953, 1961), Deaton and Muellbauer (1980), and Hausman, Leonard, and Zona (1994), where the weak separability assumption in the utility function blocks most cross-group interactions, except via the top budget.

Consideration sets, first established in Wright and Barbour (1977), are a relaxation of the assumption that individuals consider all feasible choices. They can be defined as the restricted set of shopping baskets to which a consumer pays attention in their choice process. Many reasons have been given as to why the choice set (the set of available choices) may sometimes differ from a consumer's consideration set: behavioural heuristics (Hauser, 2014), cognitive limitations (Simon, 1957), bounded rationality and seller persuasion

(Eliaz and Spiegler, 2011), imperfect information (Sovinsky, 2008), search (Caplin, Dean, and Martin, 2011), or evaluation costs broadly speaking (Hauser and Wernerfelt, 1990). Consideration sets have most recently been applied to discrete-choice models (Matějka and MacKay, 2015) and in models of revealed preference (Abaluck and Adams, 2021). Another recent contribution, Crawford, Griffith, and Iaria (2021) surveys how assumptions over the structure of consideration sets can impact demand estimation.

One concern the present Chapter speaks to is the adjustment of consideration sets to settings where it is probable that consumers consider goods both individually and in combination (e.g. Manchanda, Ansari, and Gupta, 1999; Fox and Lazzati, 2017; Allen and Rehbeck, 2022). Iaria and Wang (2019) develop a structural demand model to analyse how consumers choose product bundles, allowing for both complementarity and substitutability between goods. Their mixed logit framework incorporates a novel demand inversion technique that reduces some of the aforementioned computational complexity. Applying their model to the same data and number of brands as in Nevo (2001), they find strong evidence of demand synergies across products, driving joint purchases within the same transaction. They demonstrate that ignoring these interactions leads to biased estimates and misleading policy conclusions. Synergies were first discussed in Gentzkow (2007). Ershov, Laliberté, Marcoux, and Orr (2024) makes a similar methodological contribution also built on a random-coefficients logit framework. The authors develop a Generalised Method of Moments (GMM) estimator that uses an aggregation structure based on preference restrictions to combine aggregate sales data with specific choice probabilities per unique combination of soda, chips, or soda and chips. This results in a contraction mapping for the three aggregate "inside" goods. They address price and characteristic endogeneity using standard instrumental variables and a set of novel shift-share instruments for product variety. In an empirical review of the U.S. chips and soda market using scanner data, the authors are able to study an average of over 1300 soda and 800 chip varieties across all market-year pairs. They find large cross-category complementarities, suggesting that mergers between multi-product multi-category firms lead to lower price increases and even price decreases compared to a substitution-only model. However, this approach allows for only a small number of product categories (two in their case).

My semi-parametric demand estimation method is closest to that of Pinkse and Slade (2004). Their "spatial differentiation" approach proxies the latent Slutsky matrix of the demand system via series expansion of a distance-based measure of product differentiation. By semi-parameterising the cross-price effects, sufficient variation is introduced in the model for identification - with further heterogeneous own-price effects via interactions with covariates. This structure effectively regularises the system and captures latent complementarity and substitutability patterns that (1) help de-noise and disentangle otherwise collinear price effects, and (2) eliminate dimensionality concerns.

Lewbel and Nesheim (2019) is another paper closely related to this Chapter's approach. It makes use of a linear demand model and detailed household purchase data to separately model their choice of up to 5 fruits from a set of 27 alternatives. Our setup applies in trickier conditions, both due to requiring market-level data only, but also because corner solutions occur at the basket level, rather than over product quantities. Lewbel and Nesheim (2019) also does not directly address price endogeneity.

My proposed approach is behaviourally and data richer, as well as more computationally robust. Besides the inclusion of consideration set constraints, to minimise collinearity which arises from the Pinkse and Slade (2004) approach, the series expansion terms interacted with prices and the covariates are orthogonalised via principal components analysis (PCA). This helps reduce the number of explanatory variables (minimising overfitting) and does away with collinearity concerns entirely, minimising bias. The distance-based measure of product differentiation in this Chapter - a statistical model of co-purchase frequency first proposed in Tian et al. (2021) - is directly derived from observed purchase behaviour, allowing for greater granularity of substitute and complementary relationships between goods as understood by consumers.

Distance-based measures of product differentiation themselves are rooted in the earliest papers in the theory of industrial organisation. Lancaster (1966)'s product characteristics model, the Hotelling (1929) line, or the Salop (1979) circle are all core papers in the field, which view product differentiation as a matter of location in an abstract product space. More recently, Hoberg and Phillips (2016) have popularised cosine similarity, taking advantage of text-based measures of product similarity to identify competitors. They target a subset of firms/goods, define them over a set of descriptives identified via textual analysis of U.S. public firms' 10-K filings, and compute the cosine similarity between these sets. They assume this cosine similarity approximates the similarity between the goods' real attributes; this approach is thereby said to situate these goods in a multi-dimensional product space. In this space, the cosine similarity defines the angle between two goods with said angle assumed to negatively correlate with degree of competitive pressure. Readers are encouraged to see Ushchev and Zenou (2018), which generalise the theory behind this approach. Distance-based measures are also popular sources of instrumental variables (Gandhi and Houde, 2019).

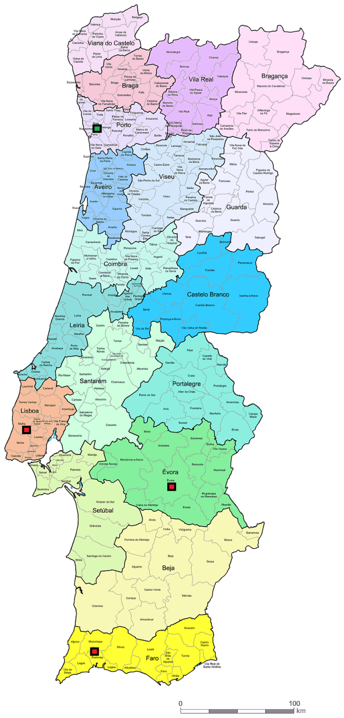
Other papers have recently experimented with semi-parametric estimation of price effects, similarly conceptualising product differentiation via distance-based measures and applying these empirically (e.g. Magnolfi, McClure, and Sorensen, 2025; Pellegrino, 2025). For these papers, the main issue is the data required to produce the cross-price effects matrix. Pellegrino (2025)'s approach targets publicly listed, (near) single-product firms, otherwise requiring (directly or indirectly) substantial data on product characteristics ; and Magnolfi et al. (2025) presents methodological difficulties in survey scaling. In using widely available co-purchase data, these limitations can be surpassed.

2.3 Data

Point-of-sale data was obtained from *SONAE MC*, a leading supermarket chain in the Portuguese food retail market. Information on transactions was collected through the stores' scanners in the moment of the sale. The dataset represents roughly 5% of all transactions conducted across four supermarkets between February 2020 to February 2023. It covers date and location of each transaction; the set of products purchased per transaction both at the product level and across 4 different product category layers; and units purchased, gross sales value, and discounts issued per product per transaction.

The four supermarkets - Supermarket Maia, Hypermarket Mafra, Hypermarket Évora, and Hypermarket Portimão - were selected prior to the data analyses out of a list of 353 shops given their geographical distribution, distance from competitors, and size heterogeneity. The selection along these lines was made to match the model's assumptions - the primacy of intra-supermarket competition in particular - while including sufficient heterogeneity to consider the internal validity of the model's outcomes under different conditions. An evaluation of supermarket participation rates - i.e. the transaction volumes observed each period per supermarket - suggests that, for our sample of grocery stores, inter-supermarket competition is in fact not a major concern (details are in the Appendix).

Figure 2.1: Shop locations



Notes: Hypermarket (red), supermarket (green). Hypermarkets are major retail hubs, meant to attract consumers to a distinct location within the urban fabric. Supermarkets are mid-sized stores targeting proximity to consumers.

Table 2.1: Transaction statistics across products and categories

	Products	**Categories**
A: Transaction size (number of items)		
Mean	60.43	5.12
Standard deviation	67.99	3.69
B: Percentile transaction size		
25th	16	2
Median	36	4
75th	78	7
C: Transaction share by size class		
$\leq$ 20 unique items / $\leq$ 2 groups	32.97%	31.04%
21-80 unique items / 3-8 groups	42.80%	50.32%
81-140 unique items / 9-14 groups	13.74%	16.97%
>140 unique items / >14 groups	10.49%	1.67%
D: Revenue share by transaction size class		
$\leq$ 20 unique items / $\leq$ 2 groups	8.04%	8.17%
21-80 unique items / 3-8 groups	30.52%	40.91%
81-140 unique items / 9-14 groups	23.21%	42.56%
>140 unique items / >14 groups	38.23%	8.36%

Table 2.2: Transaction statistics per store

Shops by Type	**Mean transaction size**	**SD transaction size**
A: Hypermarket		
Évora	70.85	77.90
Mafra	52.68	58.12
Portimão	60.52	68.50
B: Supermarket		
Maia	49.00	51.35

Table 2.3: Product availability in stores and the average revenue brought by products at each level of availability

	Products	**Avg Revenue**	**Categories**	**Avg Revenue**
At only 1 store	2%	€ 2 206.4	5%	€ 1 334.0
At 2 stores	3%	€ 693.5	-	-
At 3 stores	19%	€ 487.3	8%	€ 1 308.0
At all stores	76%	€ 970.8	86%	€ 644 261.5

Table 2.4: Shop statistics

	#	**% Unique**	**% Revenue**	**% Unique Revenue**
A: Hypermarket				
Évora	24 069	0.41%	31.91%	0.56%
Mafra	23 896	0.11%	13.73%	0.04%
Portimão	24 564	0.97%	44.00%	0.44%
B: Supermarket				
Maia	19 442	0.33%	10.36%	0.11%

Notes: The stores share 18910 products.

The following procedures were applied to filter outliers and products for which insufficient data was made available. Any products with (1) fewer than 100 transactions, (2) sold at zero price, and/or (3) without any price variation were removed. Furthermore, product without close competitors in a given month across all supermarkets were also discarded. Lastly, any products which went unobserved across a given time period in a given supermarket were assumed to not have been available for purchase.

Table 2.5: Product category statistics

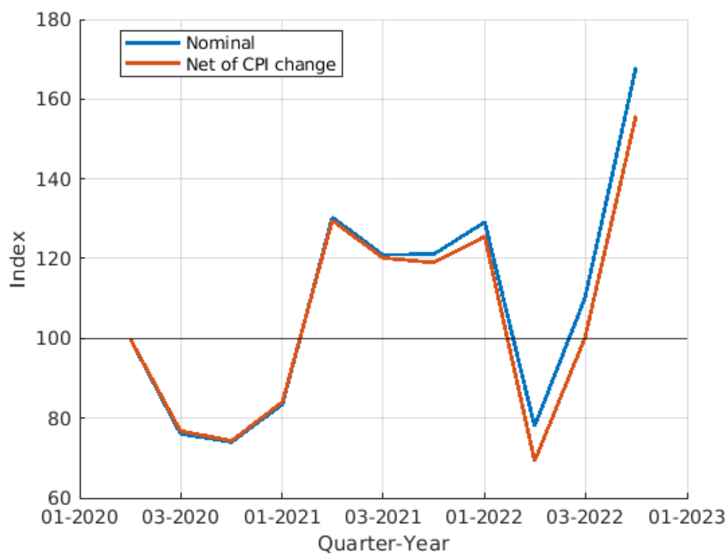
Categories	**Products**	**Store presence**	**Sales share**
Groceries (Salty)	1 766	4	4.94%
Groceries (Sweet)	2 228	4	4.64%
Soft Drinks	1 267	4	7.09%
Crates	9	4	0.01%
Hygiene	1 597	4	4.21%
Home Cleaning	2 173	4	7.18%
Frozen Goods	758	4	3.27%
Dairy and Derivates	1 478	4	7.90%
Beauty	1 665	4	2.94%
Essential Goods	690	4	3.24%
Butchery	452	4	6.35%
Fishmonger	558	4	6.47%
Charcuterie and Cheese	1 182	4	6.85%
Breakfast	1 564	4	5.55%
Fruits and Vegetables	1 037	4	10.29%
Bakery	999	4	4.35%
Wine and Spirits	1 019	4	4.83%
Take Away	468	4	2.30%
Luggage and Sport	1	4	0.00%
Bio and Healthy	6	4	0.00%
Newsagents	884	4	0.96%
DIY	744	4	1.14%
Petfood and Care	798	4	2.15%
Home Goods	1 250	4	1.75%
Stationery/Marketing	80	4	0.18%
Women Apparel	1	3	0.00%
Adult Non-Apparel	2	1	0.00%
Nursery	13	4	0.01%
Large Domestic Goods	1	3	0.01%
Small Domestic Goods	12	4	0.03%
Others	223	4	1.32%
TOTAL	24 925		

These corrections left us with 24 925 goods remaining, spanning 37 major categories, 118 sub-categories, and 571 sub-sub-categories. Each product is observed on average 1 210 times over the sample, with 175, 369, and 937 observations at the 25th, median, and 75th percentiles, respectively.

We observe 456 439 unique transactions - defined at the unit-per-product level - out of 499 047 baskets in total. The average number of observations per unique basket is 1.093; 2.16% of unique baskets are observed more than once, accounting for 10.51% of all baskets. Consistent with the shop-level results, hypermarkets exhibit larger and more dispersed basket sizes, stock more staple and unique items, and command a disproportionate revenue share. Aggregate gross sales equal € 82 636 508. Discounts total € 156 601, representing 0.19% of gross sales.

The revenue-weighted price index over the sample period is provided in Figure 2.2, both in nominal and CPI-adjusted real terms.

Figure 2.2: Revenue-weighted mean prices



The substantial observed volatility tracks the timings of the four COVID waves in

Portugal, with the price peaks generally following a given wave: the first starting in Q2 2020; the second in Q4 2020; the third in Q4 2021, and the fourth in Q2 2022.[2] Sample volatility is significantly greater than that of the CPI, the latter with only a late increase driven by energy prices and supply chain disruption, with global container freight indices - a benchmark for the average cost of shipping containers across major world trade routes - peaking at the tail-end of 2021[3] and crude oil prices peaking during 2022.[4]

In this sample, 45.8% of all goods are not purchased standalone. In what follows, I assume that this matches the structure of the true, latent, consideration set of the set of shoppers in the sample.

The aggregate consideration set is found to be full row rank. In other words, the application's setting reflects that where, within the NN-LF conditions, vector space restrictions do not apply, only non-negativity. As discussed in Chapter 1, this has two implications: all concerns raised by consideration sets and choice over combinations of goods will pertain primarily to the empirical implementation, rather than aforementioned theoretical implications; and aggregate preference parameters will be uniquely defined. Nonetheless, the Appendix includes a brief discussion on how to handle less-than-full-row-rank consideration sets in empirical specifications where it may be necessary.

2.4 Empirical Application

2.4.1 Empirical Specification

In this section, I address two concerns raised by Gandhi and Nevo (2021) regarding linear demand estimation. The most common issue is a curse of dimensionality; in these

[2]This information was obtained from the WHO COVID dashboard at data.who.int.
[3]See e.g. Monthly composite China Ningbo Container Freight Index at statista.com.
[4]This information was obtained from statista.com.

models, as the number of alternatives grows larger, the number of own- and cross-price effect parameters grows (proportional to) quadratically in these. The second concern pertains to price endogeneity. Linear demand incorporates many prices beyond a product's own, so overcoming price endogeneity concerns which arise from the simultaneity of demand and supply functions requires many exogenous and non-collinear instruments to separately identify price effects. A less discussed issue with linear aggregate demand estimation is the adjustment for individual constraints and corner solutions. In fact, representative consumer models tend to proceed without accounting for even that representative consumer's non-negativity constraint. As I show below, these introduce their own unique source of price endogeneity.

For simplicity, I will address these issues in turn, starting from the unconstrained consumer choice model. Demand is modelled at the average transaction by a representative consumer per supermarket: $\boldsymbol{Q}_{mt}/T_{mt} = \boldsymbol{q}_{mt}$. Let product utility have the following linear structure in product characteristics: $\boldsymbol{\delta} = X\boldsymbol{\beta} + \boldsymbol{v}$, for $X = [\boldsymbol{x}_1 \dots \boldsymbol{x}_K]$ a N by K matrix whose elements determine how much of a given observable characteristic k good n has; $\boldsymbol{\beta}$ a vector of the representative consumer's preference weights towards each of K observable characteristics; and $\boldsymbol{v}$ an independent, identically distributed, mean zero, unit variance, unobserved error pertaining to demand shocks and latent product characteristics. For a given supermarket m at a quarter-year pair t, unconstrained demand $\boldsymbol{q}^{\dagger}$ is therefore as follows:

$$\boldsymbol{q}^{\dagger}_{mt} = M_m^{-1}(X_m\boldsymbol{\beta} + \phi\boldsymbol{p}_{mt}) + \boldsymbol{u}_{mt} \tag{2.1}$$

for $\boldsymbol{u} = \boldsymbol{\varepsilon} + M^{-1}\boldsymbol{v}$, $\boldsymbol{\varepsilon}$ referring to the regression error. Matrix M is unknown.

The first step is to tackle the curse of dimensionality. To achieve this, I simplify this expression further by defining a Neumann series approximation of the inverse of the latent Hessian matrix on a proxy of observables W. The series approximation is helpful both for model fitting but also to minimise numerical instability and computation concerns stemming from inversion.

$$M = I - \alpha W \quad \Rightarrow \quad M^{-1} \approx g(W) = \sum_{l=0}^{L} \alpha_l W^l \tag{2.2}$$

for L the number of series expansion terms. The scaling parameter α is made to allow differences between polynomials, i.e., $\alpha_l \neq \alpha^l, \ \forall l$.

To proxy for M, I take advantage of the frequency with which products are considered for joint purchase. In the economics literature, this is an approach closest to that in Atalay et al. (2023) and with roots in stochastic choice (e.g. Manzini, Mariotti, and Ülkü, 2019). The proxy itself is drawn from Tian et al. (2021), who approach the matter of cross-price effect estimation from a network science perspective. The authors use point-of-sale data and estimate complementarity and substitution relationships by the frequency with which goods are jointly purchased, relative to what would otherwise be expected.

A modified approach to that used in Tian et al. (2021) follows. Complements are defined as goods jointly purchased more often than expected, with the degree correlated to the extent of this. Substitutes are goods that both share the same complements and are jointly purchased less often than expected, with the degree measured similarly. The sales data is converted into a product-transactions matrix S with every (row) good matched to every (column) transaction it appears in. Note that "transactions" by Tian et al. (2021) refers to non-unique "baskets". Earlier we have expressed the size of the set of all unique baskets as J; let us set T as the total number of observed transactions. From matrix S, I compute two similarity measures.

The first similarity matrix - equivalent to the *original measure* in Tian et al. (2021) - represents how similar the purchase patterns are between two products, yielding a measure of complementarity:

$$W^{(c)} = A^{(c)} \circ \cos(\theta_c) \tag{2.3}$$

The term $\cos(\theta_c)$ is a matrix whose elements are defined as, for a pair of goods $a, b \in \mathcal{K}$:

$$\cos(\theta_c)_{ab} = \frac{\Xi_{ab}}{\sqrt{\Xi_{aa}}\sqrt{\Xi_{bb}}} \tag{2.4}$$

for $\Xi = (D^{\mathcal{P}})^{-1} S (D^{\mathcal{T}})^{-1} S' (D^{\mathcal{P}})^{-1}$, $D^{\mathcal{P}}$ a diagonal matrix of the product degrees (how many transactions per good), and $D^{\mathcal{T}}$ a diagonal matrix of the transaction degrees (how many goods per transaction) . $D^{\mathcal{P}}$ normalises the product vectors, whereas $D^{\mathcal{T}}$ weighs the relative importance of each transaction as the inverse of transaction size - smaller shopping baskets in the consideration set are assumed to pertain to stronger ties between the included goods. The matrix Ξ is therefore a weighted version of the product-transactions matrix S.

The matrix $\cos(\theta_c)$ is defined under the assumption that all goods are complementary. It must therefore be supplemented by an analysis of joint purchases to determine the degree to which their frequency is *significantly* different from what would be expected of independent goods. To achieve this, we produce $A^{(c)}$, the matrix of potential complements, with elements equal to one in case of significance.

Significance is assessed under a statistical null model - the Bipartite Configuration Model (BiCM) in this case - whose goal is to create a randomised product-transactions matrix that preserves the observed degree sequence (i.e. the number of connections) for both products and transactions, while randomising all other structural properties. The BiCM is favoured in our setting as it takes the bipartite feature of the product-transactions matrix into account.

In the literature dedicated to incorporating combinations of goods (not just individual goods) as options within a discrete-choice setting, the key concern lies in separating co-purchases which result from preferences vs taste correlation. Correlated purchases can be the outcome of things like minimising trip-related costs or time and stock considerations. Preferences toward co-purchases meanwhile may come from a love for variety or complementarities derived from various setting mentioned above. Both the equation for $\cos(\theta_c)$ and the BiCM are meant to adjust for product popularity and transaction size precisely to minimise the influence of correlations.

For each product pair a and b the expected number of common transactions μ_{ab} is computed as:

$$\mu_{ab} = \frac{d_a^{\mathcal{P}} d_b^{\mathcal{P}} (\frac{1}{T}\sum_{t=1}^{T}(d_t^{\mathcal{T}})^2 - \frac{1}{T}\sum_{t=1}^{T} d_t^{\mathcal{T}})}{\frac{1}{K}(\sum_{t=1}^{T} d_t^{\mathcal{P}})^2} \tag{2.5}$$

for T the number of transactions, and $d_i^{\mathcal{P}}$ the i-th diagonal element of $D^{\mathcal{P}}$. Once we have μ we can compute the observed number of joint purchases $C = SS'$. Lastly, we compare c_{ab} with μ_{ab} using the Poisson distribution, on the assumption that in a sparse setting the probability of any two products co-occurring in a transaction is small:

$$A_{ab}^{(c)} = \begin{cases} 1 \text{ if } 1 - F_{ab}(c_{ab} - 1) < \alpha_c \\ 0 \text{ otherwise} \end{cases} \tag{2.6}$$

where $F_{ab}(y)$ is the Poisson CDF with mean μ_{ab} and α_c is a chosen significance threshold.

The second similarity matrix - equivalent to the *original substitutability measure* in Tian et al. (2021) - represents how similar pairs of products' complements are, yielding a measure of substitutability:

$$W^{(s)} = I\{(A^{(c)})'A^{(c)} > 0\} \circ A^{(l)} \circ \cos(\theta_s) \tag{2.7}$$

The expression $\cos(\theta_s)$ can be straightforwardly derived from $\cos(\theta_c)$. For every given product pair:

$$\cos(\theta_s)_{ab} = \frac{(\cos(\theta_c)\cos(\theta_c)')_{ab}}{\sqrt{\sum_{k=1}^{K}\cos(\theta_c)_{ak}}\sqrt{\sum_{k=1}^{K}\cos(\theta_c)_{bk}}} \tag{2.8}$$

The matrix $A^{(l)}$ behaves similarly to $A^{(c)}$. A pair of products is flagged as having significantly less joint purchases than would be otherwise expected of independent goods if:

$$F_{ab}(c_{ab}) < \alpha_l \tag{2.9}$$

where $F_{ab}(y)$ is once again the Poisson CDF with mean μ_{ab} and α_l is a chosen significance threshold. Should this be the case, the expression $I\{(A^{(c)})'A^{(c)} > 0\}$ is meant to identify the presence of shared complementary goods, and $\cos(\theta_s)$ provides the weights.

The following graphs and tables provide details on what complementarity matrix $W^{(c)}$ and substitution matrix $W^{(s)}$ imply regarding within- and cross-category interaction effects:

Table 2.6: Proxy details - top categories by cross-category exposure

Substitutes		**Complements**	
1	Dairy and Derivates	1	Bio and Healthy
2	Charcuterie and Cheese	2	Nursery
3	Essential Goods	3	Home goods

Table 2.7: Proxy details - top categories by own-category exposure

Substitutes		**Complements**	
1	Fruit and Vegetables	1	Bio and Healthy
2	Essential Goods	2	Marketing
3	Butchery	3	Nursery

Figure 2.3: Proxy heatmap of own- and cross-category substitution - $W^{(s)}$

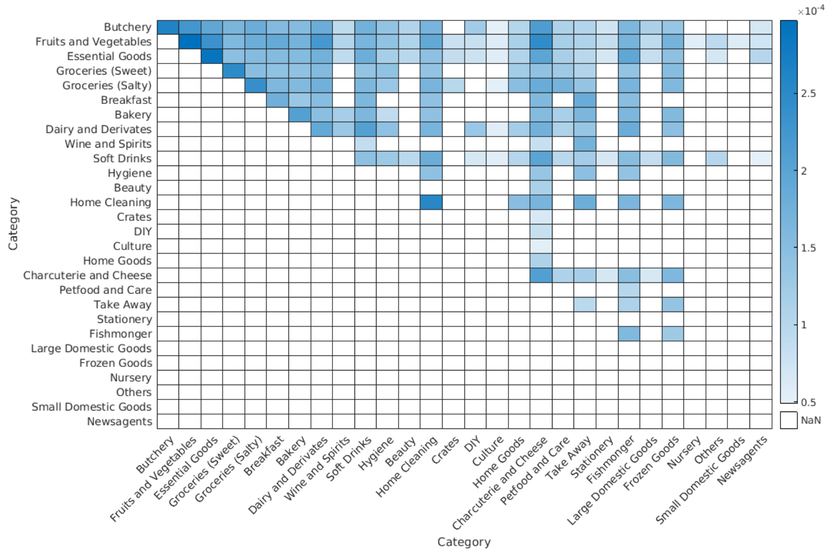
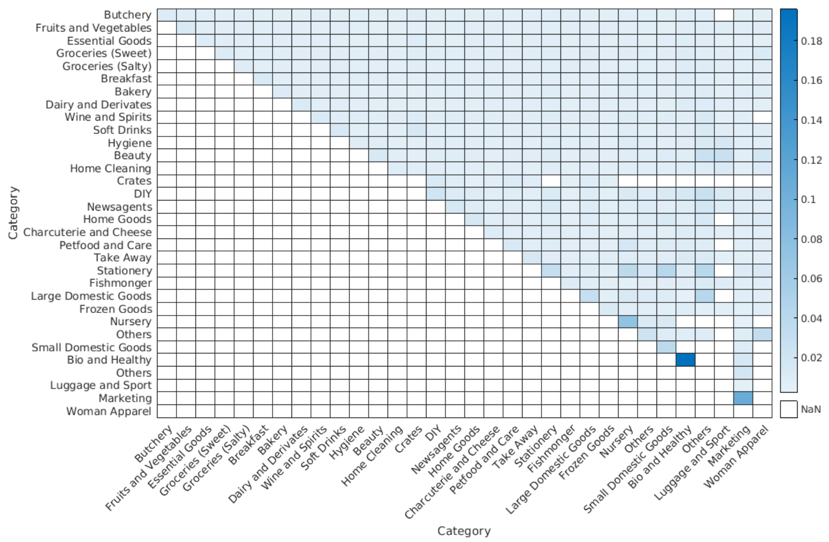


Figure 2.4: Proxy heatmap of own- and cross-category complementarity - $W^{(c)}$



The two similarity measures are tested both simultaneously and individually in the regression. Taking into account overfitting and multi-collinearity, the one providing the best fit is Tian et al. (2021)'s substitutability measure $W^{(s)}$ with two lags. All results below are provided relative to this specification.

The unconstrained linear demand model with the Neumann series approximation to the Hessian matrix inverse can be written as follows:

$$\begin{aligned} \boldsymbol{q}^{\dagger}_{mt} &\approx \sum_{l=0}^{L} \alpha_l W_m^l (X_m \boldsymbol{\beta} + \phi \boldsymbol{p}_{mt}) + \boldsymbol{u}_{mt} \\ &= X_m \boldsymbol{\beta} + \phi \boldsymbol{p}_{mt} + \sum_{l=1}^{L} \alpha_l W_m^l (X_m \boldsymbol{\beta} + \phi \boldsymbol{p}_{mt}) + \boldsymbol{u}_{mt} \end{aligned} \tag{2.10}$$

The formulation in (2.10) shows how the original regressors $X\boldsymbol{\beta}$ and $\phi\boldsymbol{p}$ are augmented by the proxy lags $W^l(X_m\boldsymbol{\beta} + \phi\boldsymbol{p}_m)$ weighted by the coefficients α_l. In practice, this means that our estimation model can be set up as a standard linear regression that includes both the original variables and their corresponding lagged versions. This is computationally favourable as it allows lags to be precomputed, and is effectively a series of linear combinations, which we may estimate via standard techniques. It also enables the separate identification of α_l from ϕ if so desired, via the delta method.

Stacking across supermarkets, and including covariate interactions with prices to account for coefficient heterogeneity, the final unconstrained model specification is as follows:

$$\boldsymbol{q}^{\dagger}_{mt} \approx X_m \boldsymbol{\beta} + (\overline{\phi} + X_m \boldsymbol{\psi}) \circ \boldsymbol{p}_{mt} + \sum_{i=1}^{M} \mathbf{1}_{\{i=m\}} \left[\sum_{l=1}^{L} \alpha_j W_m^l (X_m \boldsymbol{\beta} + (\overline{\phi} + X_m \boldsymbol{\psi}) \circ \boldsymbol{p}_{mt}) \right] + \boldsymbol{u}_{mt} \tag{2.11}$$

The second step is to address price endogeneity and instrument collinearity. The proxy lags relieve much of the latter concern, as they allow just a single instrumental variable sufficient variation to cover all prices: $W_m^l(\overline{\phi} + X_m\psi) \circ \boldsymbol{p}_{mt}^{IV}$. However, a separate concern results: subsequent proxy lag interactions with covariates $(W_m^l X_m)$ contain less new information, re-introducing multi-collinearity. To address this, a Principal Component Analysis (PCA) is applied to X to orthogonalise covariates and extract the key factors driving variation in the data. Covariates are restricted to the smallest combination of components which contains 90% of X variation or higher. The top 10 principal components in variation are retained to form heterogeneous price coefficients. While somewhat less common in Economics, the use of PCA to address multi-collinearity is commonplace in analyses of multivariate and high-dimensional data (Koch, 2013).

The last step is the handling of the Lagrangean constraints on $\boldsymbol{z}$. Consider the general, constrained form of (2.1):

$$\boldsymbol{q} = M^{-1}(\boldsymbol{\delta} + \phi \cdot \boldsymbol{p}) + A(A'MA)^{+}(\boldsymbol{\Lambda}(\boldsymbol{\delta}, \boldsymbol{p}) + \overline{\boldsymbol{\gamma}}(\boldsymbol{\delta}, \boldsymbol{p})) \tag{2.12}$$

The term $A(A'MA)^{+}(\boldsymbol{\Lambda}+\overline{\boldsymbol{\gamma}}) \neq 0$ if, for any individual within the consumer population $\exists \hat{z}_{ij} \; : \; \hat{z}_{ij} = 0$, i.e. if there are corner solutions either due to prices ($\boldsymbol{\Lambda}$) or due to consideration set heterogeneity ($\boldsymbol{\gamma}$). Together, they create a wedge between $\boldsymbol{q}$ and the unconstrained model $\boldsymbol{q}^{\dagger}$: $\Delta\boldsymbol{q} = \boldsymbol{q} - \boldsymbol{q}^{\dagger} = A(A'MA)^{+}(\boldsymbol{\Lambda} + \overline{\boldsymbol{\gamma}})$. For demand estimation, this is problematic. Because the wedge is price-dependent, it introduces an additional form of endogeneity. Not only are $\boldsymbol{\Lambda}$ and $\overline{\boldsymbol{\gamma}}$ unobserved, but we also cannot instrument said endogeneity away - the wedge is a function of price, rather than being correlated with it.

A few approaches to the non-negativity problem and $\boldsymbol{\Lambda}$ have been suggested in the literature. The most common are censored regressions, such as Tobit models (e.g. Wales and Woodland, 1983). Such an approach is inconsistent with utility maximisation, since non-negativity is introduced ex-post rather than assumed to be considered by consumers during utility maximisation. As far as I am aware, the economics literature does not discuss the matter of $\overline{\boldsymbol{\gamma}}$ and the need to account for corner solutions driven by consideration set heterogeneity at all. In our favour, however, is the representative consumer rationalisation argument in Chapter 1. There, we proved that there are an infinite number of alternative $\boldsymbol{\Lambda}$ and $\overline{\boldsymbol{\gamma}}$ that can satisfy the KKT conditions on the aggregated FOCs of the consumer population as if it were that of a representative consumer. This means we need only find the one $\boldsymbol{z}$ that delivers the best $A(A'MA)^{+}(\boldsymbol{\Lambda} + \overline{\boldsymbol{\gamma}})$ fit to the model.

I propose the following approach. First, compute $\hat{\boldsymbol{q}}^{\dagger}$, the fitted values of an initial, "naive", as-if-unconstrained regression and estimate $\hat{M^{-1}}$. Second, find $\boldsymbol{r} = \hat{\boldsymbol{q}}^{\dagger} - A \cdot \arg\min_{\boldsymbol{z}\geq 0}\{||A\boldsymbol{z} - \hat{\boldsymbol{q}}^{\dagger}||^2_{\hat{M}}\}$. Third, after getting $\boldsymbol{r}$, plug the result back onto the LHS of the regression so the dependent variable becomes $\boldsymbol{q}^{observed} + \boldsymbol{r}$ (or $\boldsymbol{q} + \boldsymbol{r}/\mathrm{sd}_Q$ in standardised units), i.e., a control-function step. Lastly, re-run the regression, obtaining a new estimate for $\hat{M}$ and repeating the process again until convergence.

I considered two alternative specifications: $\boldsymbol{r} = \hat{\boldsymbol{q}}^{\dagger} - A \cdot \arg\min_{\boldsymbol{z}\geq 0}\{||A\boldsymbol{z} - \hat{\boldsymbol{q}}^{\dagger}||^2_{\hat{M}} + \tau\boldsymbol{1}'\boldsymbol{z}\}$ and $\boldsymbol{r} = \hat{\boldsymbol{q}}^{\dagger} - A \cdot \arg\min_{\boldsymbol{z}\geq 0}\{||A\boldsymbol{z} - \hat{\boldsymbol{q}}^{\dagger}||^2_{\hat{M}} + \tau||\boldsymbol{z}||\}$. Both follow the spirit of the $\overline{\boldsymbol{\gamma}}$, either by centralising demand on a few popular shopping baskets or by decentralising it to reflect consideration set heterogeneity. Likely due to the dimension of A, our results are robust to either specification. In other words, the primary consideration is the adjustment of the empirical model to the aggregated individual corner solutions which result from prices.

2.4.2 Identification Strategy

The estimation procedure itself is carried out via Two-Stage Least Squares. I take advantage of the stream of observations on product purchases across four geographically distinct supermarkets each quarter of each year in our database, as well as the repeated cross-sectional nature of our data and our proxy lags, to define matrix Z of price instruments. I compute the average price of a given product's competitors across other supermarkets as our primary price instrument (Hausman, 1996). As exogenous variables, also in X, I include intercepts; supermarket and quarter FEs; private-label brand FEs; product category FEs at three levels of detail; and the number of supermarkets in our sample that each good is present in every time period. Lastly, following Pinkse and Slade (2004), for every additional proxy lag included in the model, additional instruments $W^j Z$ are computed.

Our estimation procedure will be successful provided $\mathbb{E}(Z'(\boldsymbol{\varepsilon}_{mt} + M^{-1}\boldsymbol{v})) = 0$.[5] Insofar as $\boldsymbol{v}$ is mean zero, independent across agents, and is composed of short-term product utility shocks, it is unlikely to play a significant role - i.e. the assumption is then equivalent to the usual exogeneity condition on IVs. This would be less so the case if it were instead composed of latent product characteristics relevant to product competition. For the latter case, the rest of the Chapter operates under the assumption that such characteristics do not shift in a competitively significant way over our period of analysis. While it is too soon to tell, early indicators seem favourable (see, e.g. Lee, 2024).

However, note that if $\boldsymbol{u} = M^{-1}\boldsymbol{v} + \boldsymbol{\varepsilon}$, then $Var(\boldsymbol{u}) = \sigma^2 M^{-2}$. Even if $\mathbb{E}(Z'(M^{-1}\boldsymbol{v})) = 0$ and our estimates are unbiased, this variance has significant effects on the error terms which need to be accounted for separately. The problem is that we do not observe M^{-1}, but rather an estimated $\hat{M}^{-1}$. Therefore, while correction may be desirable, the use of $\hat{M}^{-1}$ may compound sampling error without much benefit. There is no guarantee the same proxy that is good for predicting cross-price effects in the conditional mean is also good for capturing the covariance of latent shocks. Instead, I cluster standard errors at the store-quarter level, allowing arbitrary covariance across all product observations in that store-quarter. That is exactly what we want if we suspect the residual has a complicated cross-product structure but also want to avoid over-committing to a particular parametric structure.

2.5 Results

The results of our empirical strategy are presented below. The first-stage results pertaining to the excluded instrumental variables are as follows:

[5] That is, assuming that the corner solution wedge is successfully accounted for in the NNLS step above.

Table 2.8: Standardised 2SLS first stage

	Coefficient	SE
$PC\{IV,1\}$	0.0494*	0.0273
$PC\{IV,2\}$	−0.0015**	0.0007
$PC\{IV,3\}$	0.0062	0.0276
$PC\{X,1\} \circ PC\{IV,1\}$	−0.0001**	0.0000
$PC\{X,2\} \circ PC\{IV,1\}$	−0.0003***	0.0001
$PC\{X,3\} \circ PC\{IV,1\}$	0.0002**	0.0001
$PC\{X,4\} \circ PC\{IV,1\}$	−0.0004***	0.0001
$PC\{X,5\} \circ PC\{IV,1\}$	−0.0004***	0.0001
$PC\{X,6\} \circ PC\{IV,1\}$	−0.0002*	0.0001
$PC\{X,7\} \circ PC\{IV,1\}$	0.0031***	0.0010
$PC\{X,8\} \circ PC\{IV,1\}$	0.0005***	0.0001
$PC\{X,9\} \circ PC\{IV,1\}$	0.0005***	0.0001
$PC\{X,10\} \circ PC\{IV,1\}$	0.0000	0.0001
$PC\{X,1\} \circ PC\{IV,2\}$	0.0001	0.0001
$PC\{X,2\} \circ PC\{IV,2\}$	0.0001	0.0005
$PC\{X,3\} \circ PC\{IV,2\}$	0.0005	0.0005
$PC\{X,4\} \circ PC\{IV,2\}$	−0.0001	0.0001
$PC\{X,5\} \circ PC\{IV,2\}$	−0.0003	0.0003
$PC\{X,6\} \circ PC\{IV,2\}$	−0.0004	0.0003
$PC\{X,7\} \circ PC\{IV,2\}$	−0.0004**	0.0002
$PC\{X,8\} \circ PC\{IV,2\}$	0.0000	0.0002
$PC\{X,9\} \circ PC\{IV,2\}$	0.0006***	0.0002
$PC\{X,10\} \circ PC\{IV,2\}$	0.0005***	0.0002
$PC\{X,1\} \circ PC\{IV,3\}$	−0.0005**	0.0002
$PC\{X,2\} \circ PC\{IV,3\}$	−0.0001	0.0001
$PC\{X,3\} \circ PC\{IV,3\}$	0.0003	0.0002
$PC\{X,4\} \circ PC\{IV,3\}$	0.0005***	0.0001
$PC\{X,5\} \circ PC\{IV,3\}$	0.0001	0.0002
$PC\{X,6\} \circ PC\{IV,3\}$	0.0002	0.0002
$PC\{X,7\} \circ PC\{IV,3\}$	−0.0060**	0.0025
$PC\{X,8\} \circ PC\{IV,3\}$	−0.0012***	0.0004
$PC\{X,9\} \circ PC\{IV,3\}$	−0.0002	0.0002
$PC\{X,10\} \circ PC\{IV,3\}$	−0.0000	0.0002

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

On the one hand, the majority of significant effects are primarily observed amongst the interaction effects between the product characteristics and the Hausman instrument without proxy lags; on the other hand, two of the effects greatest in magnitude pertain to an interaction effect including the second proxy lag. To assess instrument informativeness and relevance across both the base and consideration-set-adjusted model, I obtained the following diagnostics:

Table 2.9: Model diagnostics and PCA summary

	Base model	NN-LF-adjusted model
PCA		
# PCs for 90% variance	76	76
Variance explained by top 10 PCs (%)	76.6	76.6
Fit & diagnostics		
Adjusted R^2	0.630	0.564
F-test (122, 557514)	42799***	862650***
Durbin-Wu-Hausman χ^2 (33)	142.74***	5565.4***

Notes: Robust standard errors in parentheses. *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$. PCA is computed on the initial covariate set X. Degrees of freedom for the Wald test are shown as $(\text{df1}, \text{df2})$; the Durbin–Wu–Hausman statistic displays the χ^2 degrees of freedom in parentheses.

The PCA applied to X and its proxy lags to orthogonalise covariates and extract the key factors driving variation in the data reduces their number to 76. The top 10 principal components are used to form heterogeneous price coefficients in both specifications, and are shown to explain up to 76.6% of the variation across all covariates. Combined with quarter-year and supermarket FEs as well as the proxy lags, the total number of regressors is 122, over 557514 observations. The NN-LF-adjusted model converges after three iterations.[6]

The model retains substantial explanatory power even in its more restrictive version. The difference in fit is suggestive how much of the unadjusted results may be spurious. In each case, the included regressors are jointly informative, and the Hausman instrument - supported by its proxy lags - successfully minimises existing price endogeneity, which as

[6]Convergence is assessed on whether the largest before-and-after entrywise change between either the vector of coefficients $\boldsymbol{\theta}$ or the wedge $\boldsymbol{r}$ relative to their previous maximimum element respectively is less than a pre-specified tolerance, namely 0.1. They are 0.0681 and 0.4766 respectively after the third iteration.

expected further decreases once adjusting for corner solutions.

The coefficients pre- and post-NN adjustment are as follows, focusing on the parameters of greatest interest:

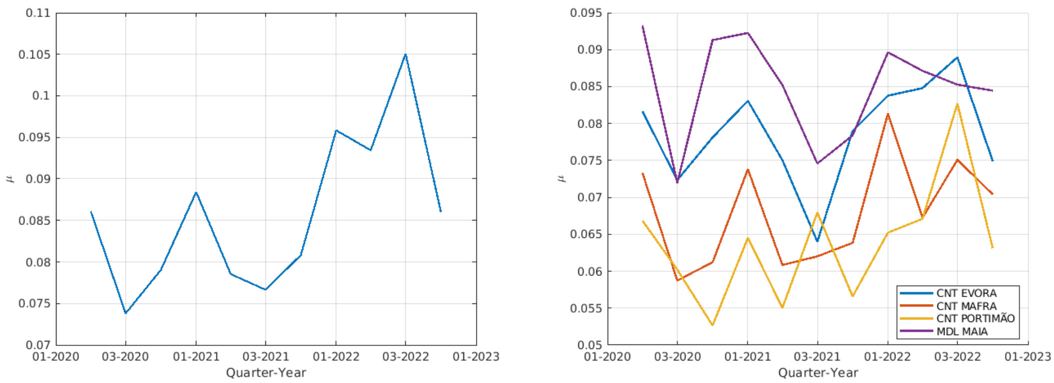
Table 2.10: Standardised 2SLS parameters of interest (Pre-NNLS vs Post-NNLS)

	Base model	NN-LF-adjusted model
$\bar{\phi}$	8.1426**	2.0504
	(3.3137)	(3.2655)
η_1	−0.0223	−0.1227***
	(0.0180)	(0.0186)
η_2	−0.1305	−0.6104***
	(0.0859)	(0.0898)
η_3	−0.0856	−0.5029***
	(0.0766)	(0.0788)
η_4	−0.0046	0.0029
	(0.0085)	(0.0085)
η_5	0.0577	0.2766***
	(0.0528)	(0.0536)
η_6	0.0106	0.1813***
	(0.0416)	(0.0429)
η_7	0.1189**	0.4580***
	(0.0510)	(0.0536)
η_8	0.0027	0.0950***
	(0.0236)	(0.0238)
η_9	−0.0515*	−0.1444***
	(0.0261)	(0.0271)
η_{10}	0.0482**	0.1510***
	(0.0213)	(0.0226)
α_1	5.3587	15.6740**
	(6.2250)	(6.4990)
α_2	−3.7802	−15.6690***
	(3.5350)	(3.6969)

Notes: Robust standard errors in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Most notably, prior to the adjustment, neither the proxy nor most of the price interaction effects are significant. After accounting for price endogeneity due to the aggregated individual corner solutions, most effects are significant and, in the case of the proxy parameters, the magnitude is particularly expressive. To analyse the implications of the consideration set constraints for the adjusted model where it comes to price elasticities, an analysis of the resulting mark-ups is presented below. The revenue-weighted mean mark-ups are presented both for the case where goods are assumed to be priced independently (product-level pricing assumption) and by their respective supermarket (retailer-level pricing assumption).

Figure 2.5: Revenue-weighted mean mark-ups - product- and retailer-level



In each case the average mark-up shows some seasonality, with an upward trajectory before returning to baseline in the last quarter of the sample. On the left, throughout the sample period, mark-ups average around 8.5% though under some volatility - perhaps less than one may othewise expect for this sample period. We can observe a distinct peak in the first quarter of 2021 (≈8.5%), and another peak in the third quarter of 2022 (≈10.5%). On the right, volatility tracks that on the left - Supermarket Maia averages a 8.5% mark-up; Hypermarket Évora at 7.5%; Hypermarket Mafra at 7%, and Hypermarket Portimão at 6.5%.

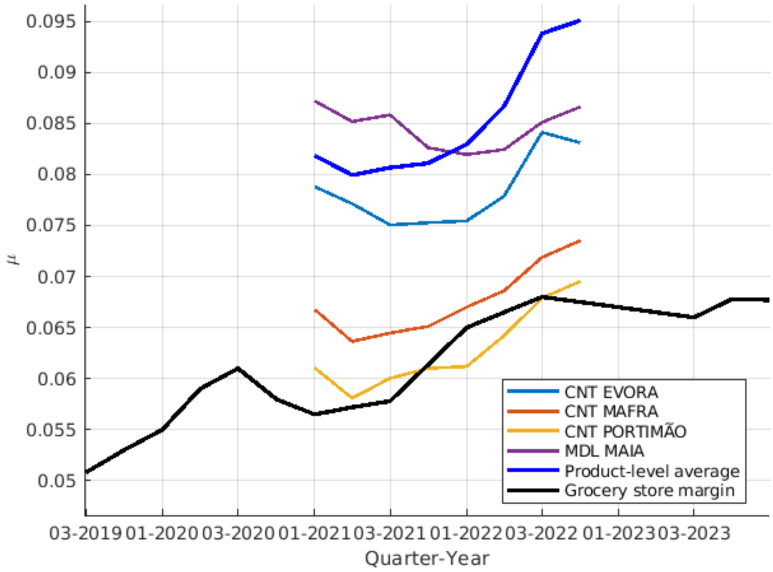
Regardless of pricing assumption, the results obtained in the present Chapter match survey estimates of United States grocery store margins remarkably well.[7] U.S. data is used here due to the unavailability of quarterly figures for Portugal.[8] Figure 2.6 provides

[7]Data from U.S. Census Bureau's Quarterly Financial Report (QFR) for Retail Trade, reproduced from a report by the White House Council of Economic Advisors, originally obtained at: https://bidenwhitehouse.archives.gov/cea/written-materials/2024/06/20/update-grocery-price-inflation-has-cooled-substantially.

[8]The data which does exist is annual, see Banco de Portugal's Quadros do Setor in https://www.bportugal.pt/QS/qsweb/Dashboards. It roughly corroborates the U.S. trends.

a side-by-side comparison of the four-quarter moving average mark-up in our sample and in the U.S. for our sample period.

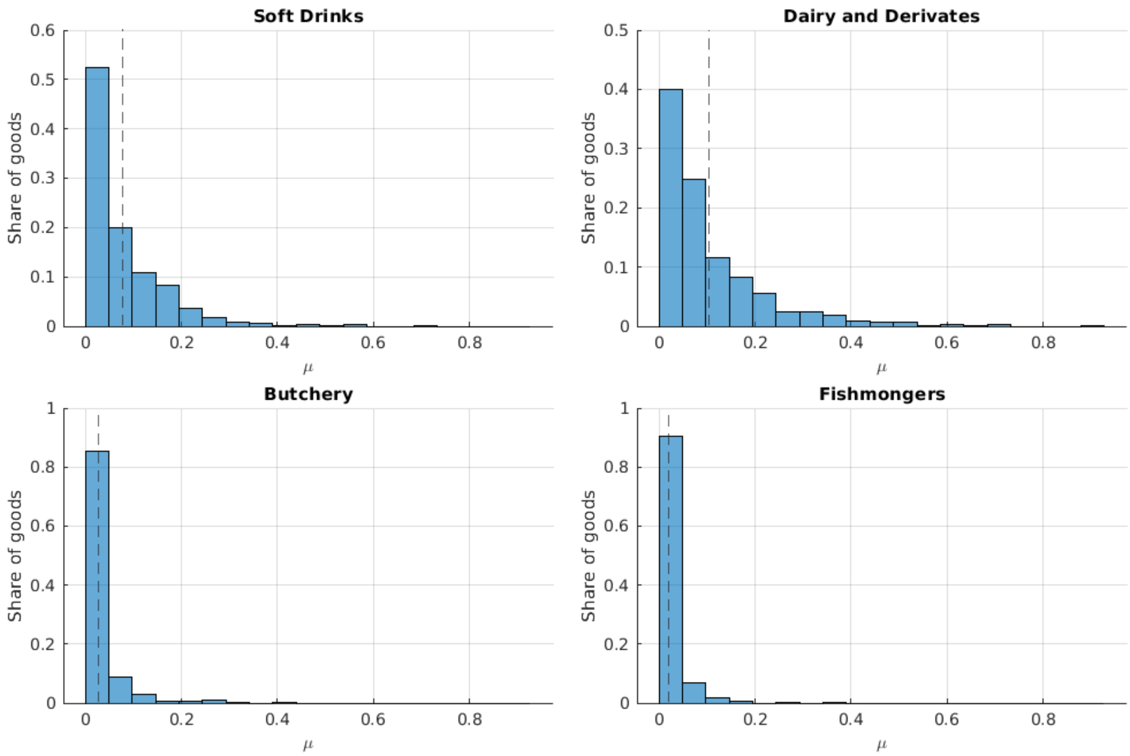
Figure 2.6: Revenue-weighted four-quarter moving average mark-ups: U.S. survey estimate and sample



The trends are near identical - the U.S. trend leading the Portuguese trend by about a quarter - though levels differ cross-country. U.S. mark-ups see a trough of 5.7% in the first quarter of 2021 and a peak of 6.8% in the third quarter of 2022: roughly half of those observed under product-level pricing, though it matches the lower bound of the supermarket mark-ups under retail-pricing. Any minor differences may nonetheless lie in grocery store margins excluding supplier mark-ups. Draganska, Klapper, and Villas-Boas (2010) find a rough 50/50 split in mark-ups between retailers and manufacturers in the German market for ground coffee, using scanner data from a national sample of stores belonging to six major retail chains. Portuguese and U.S. grocery store margins may also differ. Any differences are therefore within the expected interval for double marginalisation. Differences may also be explained by other costs accounted for in the grocery store profit margin, which are unobserved to researchers.

Revenue-weighted mark-up distributions are right-skewed, though there is some degree of heterogeneity across product categories. The following Figure provides a selection of these under the product-level pricing assumption:

Figure 2.7: Average mark-up distribution - selected categories

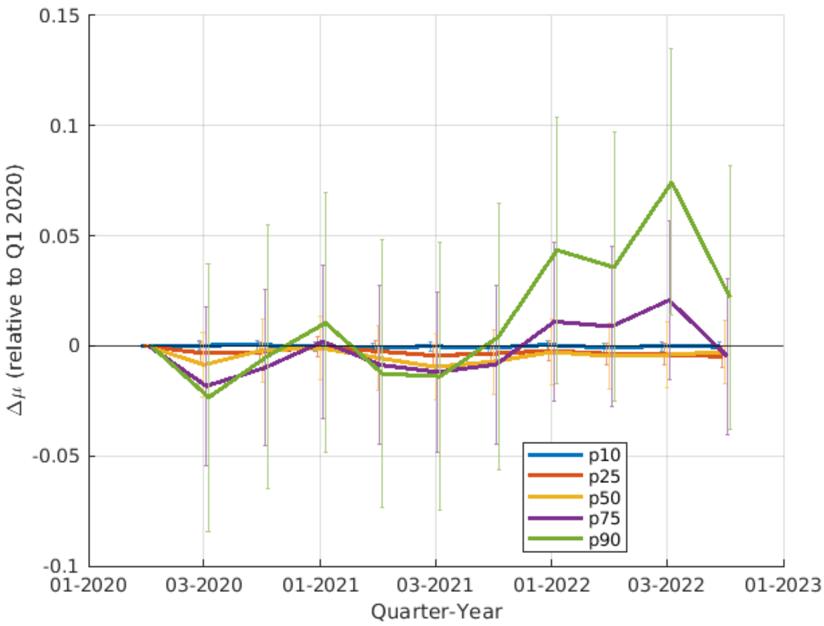


These results track some of the interaction effects revealed earlier. For example, "Butchery" ranks amongst the highest in own-category substitution exposure, while "Dairy and Derivates", despite ranking highly in own-category substitution exposure, also ranks well in own-category complementarity. "Fishmongers" rank low on own-category substitution and complementarity, but highly on cross-category substitution. "Soft Drinks" distribution is most likely connected to significant differentiation in the category, which includes "Beer", "Juice", "Water", and "Soda".

To obtain further insight on the distribution and significance of the mark-up trends,

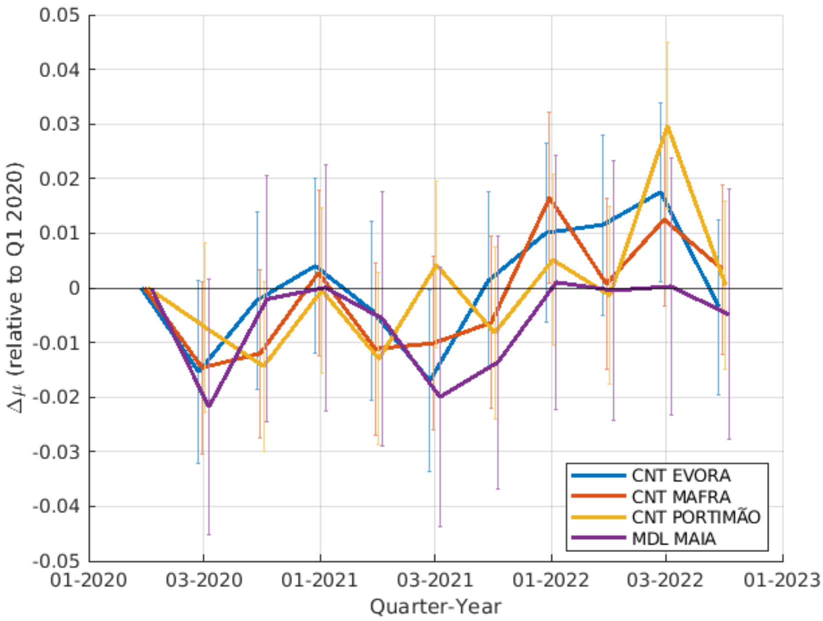
Figure 2.8 provides robust confidence intervals on the revenue-weighted mean mark-up estimates across mark-up percentiles, under the product-level pricing assumption.

Figure 2.8: Revenue-weighted mean mark-ups - percentiles



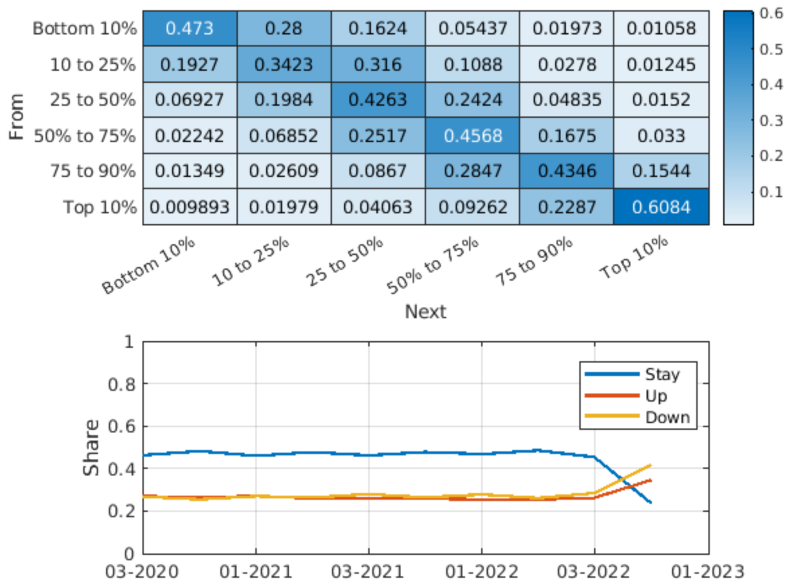
Mark-up changes remain insignificant for most goods across the sample period, except for a 7.4% increase in the third quarter of 2022 amongst the highest percentile group, which seems to start to some extent, though non-significantly, two quarters prior (3.6% and 4.3% respectively), before dissipating the next period. This corresponds to the exact period where the CPI most closely aligns with sample inflation, likely due to supply chain disruption and an international energy shock. Notably, the increase in mark-ups observed in 2022 is not driven by supermarket heterogeneity, as can be seen in Figure 2.9 (here under the retailer-level pricing assumption):

Figure 2.9: Revenue-weighted mean mark-ups - supermarkets



Across pricing specifications, the most robust trend shift throughout the sample period is the late mark-up raise around Q4 2022. To explore the causes of this apparent late jump in market power, I consider a few measures of product mobility across the mark-up distribution in Figure 2.10. First, I recompute percentile bins within each category every quarter-year period, assign each product to a bin, and track transitions across adjacent quarter-years. A transition heatmap maps cross-time transitions, while a companion line plot shows the time path of said transitions.

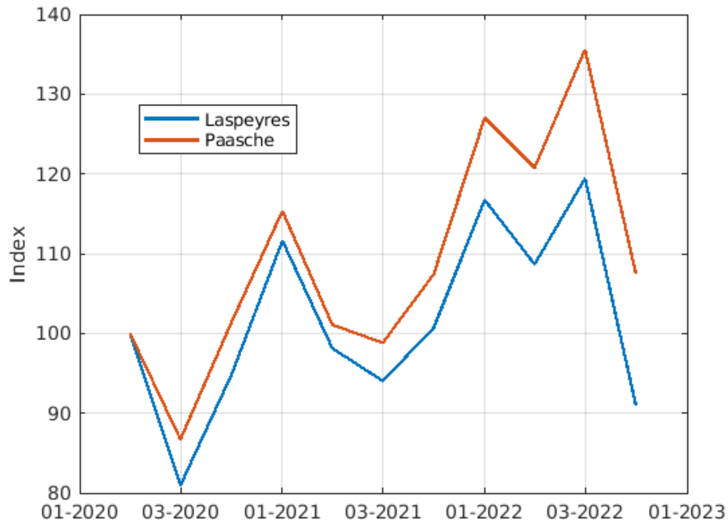
Figure 2.10: Mark-up transition dynamics and transition shares over time



Overall, we observe substantial time persistence, coupled with some degree of mean reversion. Products at the top end of the mark-up distribution, which we have noticed above to be a key driver of the overall mark-up increase observed at the end of our sample period, are especially persistent in their positioning. However, as can be seen in the line plot, persistence is shaken around the time where we have elsewhere observed a rise in market power, in mid-to-late 2022.

To explore this further, in Figure 2.11 we observe Laspeyres- and Paasche-type fixed-basket mark-up indices, based on initial and final revenue weights respectively.

Figure 2.11: Fixed-basket mark-up indices



Divergence between them diagnoses reallocation effects; and in fact, as of the third quarter of 2022, the Paasche index sat above the Laspeyres index, revealing shifting preferences. This reflects reports from industry players, who suggest two key factors as playing an important role in this change in behaviour: (i) brand switching as a result of stockouts; and (ii) a move away from in-person dining, which as of 2024 had yet to rebound.[9] A specific group of high-revenue products, specifically of a high mark-up percentile, appears to have decoupled from other goods in early 2022.

How do these results compare to the existing literature? Dopper, MacKay, Miller, and Stiebale (2024) estimates separate random coefficient mixed logit models by category and year for a panel dataset of consumer products in the US. In that paper, a mean of median category-level mark-ups - which overweights less revenue-relevant, lower-mark-up goods relative to the measure presented here, and would thus supposedly be lower than our figure - points to an average mark-up of 63% in U.S. grocery store scanner data. Additional category-specific mark-up estimates from the literature are provided below:

[9] See e.g. https://www.mckinsey.com/industries/consumer-packaged-goods/our-insights/state-of-consumer, and https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/survey-us-consumer-sentiment-during-the-coronavirus-crisis.

Table 2.11: Revenue-weighted median own-price elasticity by category

Category	This Chapter	Lit. Estimate	Citation
Beer	-44.93	-4.74	Miller and Weinberg (2017)
Breakfast cereal	-16.02	-2.42	Backus, Conlon, and Sinkinson (2021)
Yogurt	-6.99	-4.05	Hristakeva (2022)

Notes: The Miller and Weinberg (2017) estimate is the median product-level elasticity obtained with the RCNL-1 specification. The Backus et al. (2021) estimate is the median product-level elasticity obtained with the "prices only" specification. Hristakeva (2022) reports a mean product-level elasticity from 2001–2010. My corresponding estimate is the revenue-weighted median own-price elasticity, across all periods, for the subcategories: "Soft Drinks" - "Beer"; "Breakfast" - "Cereal"; and "Dairy and Derivates" - "Yogurt and Desserts", net of "Desserts".

The relative level difference may be connected to a difference between Portuguese and American tastes; relative to Portugal, per capita the US consumes roughly 32.4% more beer (Kirin Beer, 2020), roughly the same in breakfast cereal (Lopes et al., 2017; Lin et al., 2003), and 3-4 times less yogurt (Marramaque and Cardoso, 2021; Sebastian et al., 2024). The most noticeable feature of Table 2.11 is however the magnitude difference between the measures in the literature and the findings in this Chapter. A source of the magnitude difference may in how the random coefficient mixed logit ignores cross-category effects. This has predictable implications for price elasticity estimates. Consider a stylised regression parameter for own-price effects, in the case where cross-category effects are ignored (or more broadly where market boundaries are understated):

$$\hat{\beta}_i = \frac{Cov(q_i, p_i)}{Var(p_i)} = \frac{\sum_k J_{ik} Cov(p_k, p_i)}{Var(p_i)} = J_{ii} + \sum_{k \neq i} J_{ik} \frac{Cov(p_k, p_i)}{Var(p_i)} \tag{2.13}$$

for $J = \phi \cdot A(A'MA)^{+}A'$. $J_{ii} < 0$ while J_{ki} may be positive or negative depending on the product pair. Prices are strategic complements under Bertrand competition, so retailer prices co-move weakly positively across substitute goods and weakly negatively otherwise, suggesting $Cov(q_i, p_i) \geq 0$ and $Cov(q_i, p_i) \leq 0$ respectively. The cross-price terms J_{ki} behave similarly, meaning that the summation contributes positively to $\hat{\beta}_i$. Thus, if cross-category price effects are ignored (or included, as in this Chapter), own-price effect estimates are biased towards (away from, respectively) zero; this underestimates (overestimates, respectively) price elasticities, possibly producing excessive (insufficient, respectively) estimates of market power and understating (overstating, respectively) pass-throughs. This also explains why the retailer-level mark-ups appear to be on average lower, as cross-price effects are directly incorporated into the optimal pricing.

Other arguments can help reconcile the rest of the literature with the reported margins in our sample and at U.S. grocery stores. Single-category research has in the past targeted categories where American suppliers' market power is especially evident, either due to market share concentration (e.g. soda, mark-ups estimated at between 19.9 and 63.9%; Dubé, 2005) or brand proliferation by large multi-product firms (e.g. ready-to-eat cereals, mark-up estimated at 35.8% under assumption of single-product firms; Nevo, 2001). In the case of Dopper et al. (2024), the sample ends in 2019 and observes stable prices, while our sample starts in early 2020 and contains significant price volatility. In any case, it may be to some extent hard to rationalise low grocery store profit margins, such as those observed in the US Census Bureau's Quarterly Financial Report, using mark-ups as high as those which have been previously reported in the literature.

2.6 Discussion and Conclusion

This Chapter has shown that aggregated linear demand can be estimated in a high-dimensional retail setting once the main empirical obstacles are addressed directly. Building on Chapter 1, I proposed an empirical strategy that combines a consideration-set-based proxy for substitution and complementarity effects with a 2SLS specification and a control-function correction for aggregated individual corner solutions. The central contribution of the Chapter is precisely this correction: once the Lagrangean constraints on shopping-basket choice are taken seriously, the resulting adjustments separate movements in demand driven by true price effects from those driven by changes in the set of active consumers. In doing so, the approach makes it possible to recover economically meaningful price elasticity parameters across a product assortment that would be difficult to study otherwise.

The results suggest that empirical work with linear demand has been constrained less by the functional form itself than by the difficulty of estimating it credibly in large product spaces. Once dimensionality is reduced in a disciplined way, instruments are extended to cover rich price interactions, and aggregated corner solutions are explicitly incorporated, linear demand becomes a scalable and informative tool for studying competition in differentiated products markets. This is particularly valuable in settings where cross-category interactions, broad assortments, and evolving shopping baskets make conventional category-by-category approaches restrictive.

An open question is the realism of assuming that consumers agree on what goods are (closer) substitutes or complementary. By focusing on common demand slopes, the model from Chapter 1 makes a testable claim; this Chapter clears the way for the use of non-anonymised transaction data to flexibly estimate individual price effects and test Eaton and Lipsey (1989)'s observation. This is left for future work.

3

Endogenous Product Design in Aggregated Linear Demand Models

Contents

3.1 Introduction **75**

3.2 Literature Review **78**

3.3 Theoretical Framework **82**

3.3.1 Setup 82

3.3.2 Characteristics-Based Linear Demand 83

3.4 Optimal Product Design under Monopoly Pricing **89**

3.4.1 One-Attribute Case 89

3.4.2 Attribute Exclusivity 91

3.4.3 Attribute Non-Exclusivity 93

3.4.4 Discussion 98

3.5 Optimal Product Design under Single-Product Firm Pricing **99**

3.5.1 Attribute Exclusivity 100

3.5.2 Attribute Non-Exclusivity 102

3.5.3 Discussion 108

3.6 Firm- and Product-level Heterogeneity in Attribute Costs **110**

3.7 Product Entry and Market Outcomes **115**
3.8 Numerical Example **118**
3.9 Discussion and Conclusion **120**

3.1 Introduction

Consider the following setup.[1] A firm wishes to maximise profits by strategically setting the price(s) and characteristics of its good(s) within a product category. Consumers derive utility from said good(s) based on those characteristics; the characteristic choices of the firm's competitors also shape demand elasticities and the intensity of price competition. We want to know how said firm may optimise their good(s)' characteristics and what product differentiation and market structure are likely to arise as a result. How can we determine this?

I answer this research question in this Chapter. An adequate modelling of endogenous product design is of broad interest. In industry, firms may benefit from a systematic understanding of design for strategic decision-making. Policy-makers and regulators regularly consider and apply minimum standard requirements, e.g. right-to-repair laws and the the USB-C standard in consumer electronics, or minimum free allowance for cabin bags in flights. Despite a recent resurgence in these types of requirements (Zeitlin and Rangoni, 2025), adequate analytical methods to consider their implications have yet to be developed.

To formalise a preliminary theory of endogenous product design, I focus on a simple version of the model in Chapter 1, with a set of homogenous consumers rationalisable

[1] I have no relevant or material financial interests that relate to the research described in this Chapter. I gratefully acknowledge financial support by the Department of Economics at the University of Oxford and the Fundação para a Ciência e Tecnologia of the Ministry of Education, Science and Innovation (Portugal).

through a representative consumer. I introduce a novel theoretical approach for analysing firm behaviour in differentiated products markets with a finite number of goods, firms, and product characteristics. It proposes a full, characteristics-based specification for linear demand. Firms are able to set both orientation and intensity across all the characteristics of the goods they sell. Different orientations between goods imply horizontal differentiation, and different intensities imply vertical differentiation; the model can therefore speak to both dimensions. Furthermore, product characteristics are allowed to differ in their relevance to product utility or differentiation. Characteristics differ in their utility but also in their salience to product market competition, and the two need not be related: the color on a phone may be low utility but high salience; data privacy may be high utility but low salience. Lastly, the theoretical approach allows characteristics to impact substitution/complementarity between goods through unique interactions with each other.

At the heart of the proposed characteristics-based linear demand approach sits an orthonormalised representation of product characteristics, which I rename as *attributes*. Attributes provide the necessary tractability for a general model of product design. The key technical step is then a general approximation of the Hessian matrix of the representative consumer's utility function. In linear demand, the Hessian matrix, which reflects diminishing marginal utility due to satiation, is proportional to the price effects. I propose an approximation assuming that said matrix is at least partially determined by observed product characteristics. The approximation keeps exactly the structure that is identified by these, without imposing extra structure on what we do not observe.

I consider two market structures: one with a single multi-product monopolist, another with multiple single-product oligopolistic firms. I analyse each under different extents of control over their own product characteristics, as well as both where firm- and product-level characteristic costs do and do not differ, introducing asymmetries. Characteristics costs are used to reflect how increasing the intensity of a given characteristic takes engineering effort, better inputs, brand investment, testing, or foregone room for other features. Costs are quadratic to represent how each increment gets harder as the easy improvements are exhausted first.

Under monopoly pricing, I show the model has a unique equilibrium with no incentive for horizontal differentiation in observed characteristics. The monopolist prefers attribute alignment across all its goods. Cannibalisation alone is not a sufficiently strong incentive to diversify the product range; instead, alignment stops the monopolist from spreading attractiveness across different directions and instead lets it push the whole product line towards the direction the consumer likes most. The whole line is pulled toward the same attractive part of product space. The monopolist is otherwise indifferent over the extent of vertical differentiation across (however many) goods, once aggregate attribute intensity is fixed.

Under single-product-firm pricing, I prove the existence of a unique symmetric equilibrium in product characteristics, without either horizontal or vertical differentiation in observed attributes. Competing firms do however tilt their product design across more attribute. Competition makes firms more balanced in the attributes they load. To minimise business stealing by other market participants, they increasingly emphasize the attributes with lesser competitive salience as the number of competitors increases.

For a degree of asymmetry, I introduce firm- and product-level heterogeneous characteristics costs. Under monopoly pricing, asymmetry breaks the indifference in how attributes are distributed across goods. The monopoly instead loads all attributes onto a single good: the good with the lowest weighted costs. A higher value is placed on lower costs in the higher-utility attributes. This suggests that a single good is sufficient to extract surplus from the consumer, despite complementarity between goods not being ruled out ex-ante. Under single-product-firm pricing, cost asymmetry can and does sustain greater variety in orientations, and therefore horizontal differentiation. While a closed-form solution is not achieved, the model remains sufficiently nimble to converge numerically.

Lastly, the Chapter addresses demand prediction for new goods. I find that greater similarity between incumbents and entrants lowers incumbent demand and profits and discourages entry. If entry takes place, it shifts incumbent design in the exact opposite direction from the entrant's. This matches recent empirical research suggesting that niche designs are a common early entry strategy, with a greater focus on a broader audience over time (Gong, 2021).

The remainder of the Chapter is organised as follows. Section 3.2 reviews the literature on product differentiation and endogenous product design. Section 3.3 sets out the theoretical framework and develops the characteristics-based representation of linear demand used throughout the Chapter. Section 3.4 studies optimal product design under monopoly pricing, while Section 3.5 turns to the case of multiple single-product firms that set prices independently. Section 3.6 extends the analysis to firm- and product-level heterogeneity in attribute costs. Section 3.7 examines product entry and its implications for market outcomes and post-entry design incentives. Section 3.8 presents a numerical example to illustrate the model's mechanisms. Section 3.9 concludes.

3.2 Literature Review

There is a relatively long tradition in microeconomic theory of approximating product competition via some measure of closeness in a product space. Early models endogenise horizontal differentiation via location choice, be it over a line, a circle, or a spoke (Hotelling, 1929; D'Aspremont, Gabszewicz, and Thisse, 1979; Salop, 1979; Chen and Riordan, 2007). In subsequent models of vertical competition (e.g. Spence, 1976; Shaked & Sutton, 1982;

Moorthy, 1988, Motta, 1993), quality choice is the measure instead; a utility function unidimensional parameter that raises consumers' willingness to pay and (usually) marginal or fixed costs, and along which a firm locates optimally. An earlier strain of research on product variety as the optimal number of differentiated-product firms in a market is covered in some detail in Lancaster (1990).

Another tradition in the Economics literature considers distance-based measures of product differentiation in characteristics, rather than product, space. Lancaster (1966) pioneered this approach, whereby a product is defined as a vector whose dimensions reflect different abstract characteristics, and consumers have preferences defined over said characteristics, purchasing products given their respective characteristics bundle. More recently, characteristics-based models have begun to incorporate measures of competitive pressure defined as a function of the similarity between product pairs in a product characteristics space. In recent empirical microeconomics research, the most common such measure is cosine similarity. Hoberg and Phillips (2016) take advantage of text-based measures of product similarity to identify competitors. They target a subset of firms/goods, define them over a set of descriptives identified via textual analysis of 10-K fillings by US public firms, and compute the cosine similarity between these sets. They assume this cosine similarity approximates the cosine similarity between the goods' real characteristics; this approach is thereby said to situate these goods in a multi-dimensional product space. In this space, the cosine similarity defines the angle between two goods, with said angle assumed to negatively correlate with degree of competitive pressure.

More recently, Pellegrino (2025) applies this measure to mark-up estimation in a quasi-linear quadratic utility model and Cournot competition. A literature on the implication of product design for welfare based on this model has sprung up: Voelkening (2026) studies characteristic choice in a single-product duopoly setting; Miyashita (2026) extends this analysis to multiproduct monopoly.

However, cosine similarity faces issues which limit its applicability in a more generalised model. As a measure of competitive pressure, it lacks a micro-foundation, ascribes equal weights to all characteristics regardless of their importance to consumers, only accounts for observed characteristics (exogenously chosen by the researcher), and does not address the overlapping information these may contain for defining a given product. Mathematically, it also yields a poorly behaved demand Jacobian in an equilibrium where product characteristics are endogenously chosen (Voelkening, 2026). To address these issues, the present Chapter introduces a general measure of product competition based on product characteristics, which nests cosine similarity, and provides the necessary mathematical tractability to allow for multi-dimensional characteristic choice across any number of firms and goods. This echoes the Lancaster (1966) characteristics approach, adjusted so that characteristics are both the source of product utility and of consumption satiation.

A few papers have explored both horizontal and vertical differentiation simultaneously. Von Ungern-Sternberg (1988) can be understood as the first model integrating both, where horizontal differentiation is defined as distance in the Salop (1979) circle, but firms may also endogenously select their own transport costs, understood as how close a good is to be perceived as "general purpose". Economides (1993) considers a model with one horizontal characteristic and one vertical characteristic with firms also distributed over a circle. Three related papers - Perloff and Salop (1985), Wolinsky (1986), and Anderson and Renault (1999) - are seminal contributions to the product differentiation literature but focus on pricing, leaving product design - understood there as consumer-firm match value - exogenous.

More recent work allowing for endogenous design is primarily based on Johnson and Myatt (2006); the associated literature has come to be known as "targeted" product design.

This refers to targeting by the degree of tailoring to a broad versus narrow set of consumers (Bar-Isaac, Caruana, and Cunat, 2023). Johnson and Myatt (2006), building on Mussa and Rosen (1978), analyses how changes in the distribution of consumer preferences affects a firm's demand and profitability. It argues that firms find the most success by pursuing either a mass-market strategy with low dispersion or a niche strategy with high dispersion. Johnson and Myatt (2006)'s definition of a consumer's valuation for a product is a surprising match to the present Chapter's own approach - there, preferences are normally distributed, with the variance-covariance matrix a function of weights placed on different product characteristics. Insofar as quasi-linear quadratic utility can be interpreted as a mean-variance model, with the Hessian matrix the variance-covariance matrix of the demand for goods (a common interpretation in Finance, see e.g. Campbell and Viceira, 2002, Ch. 2; Koijen and Yogo, 2019), the intuition follows. However, many of the results in Johnson and Myatt (2006) rely on an assumption that firms control not just their product characteristics, but the full dispersion of preferences; results also do not fully carry over to market competition or in more complex settings with asymmetries. Bar-Isaac, Caruana, and Cunat (2012) adapt Johnson and Myatt (2006)'s analysis to a competitive environment and attach a consumer search model (see also Larson, 2013); Bar-Isaac, Caruana, and Cunat (2023) allows for marginal cost asymmetries in the same setting.

The model in the present Chapter attempts to generalise these efforts, allowing for multi-dimensional choice on both the extensive and intensive side, for any number of firms and goods. Rather than targeting consumers, the focus is on firm behaviour and strategic choices in equilibrium. Product characteristics themselves are allowed to differ in their salience to consumers, whose heterogeneity follows from the model in Chapter 1 rather than being stylised. The suitability of the present Chapter's model for empirical implementation also constitutes a unique contribution (see Appendix).

Lastly, this Chapter contributes also to the literature on asymmetric competition.

There are many sources of competitive asymmetry, for example: asymmetry in market power (as in Stackelberg competition), information (e.g. Akerlof, 1970), consumer captivity (e.g. Varian, 1980), capacity constraints (e.g. Kreps and Scheinkman, 1983), marginal costs (e.g. Blume, 2003), and most recently in data (Rhodes and Zhou, 2024) and network effects (Peitz and Sato, 2025). In this Chapter, I focus on asymmetry in product design; when it endogenously arises from market conditions; and how it reflects asymmetry in good- and firm-level marginal costs in product characteristics' intensity.

3.3 Theoretical Framework

3.3.1 Setup

In this Chapter, we focus on a simple version of the theoretical framework of Chapter 1, with a set of consumers with homogeneous preferences, modeled through a representative consumer (Gorman, 1953, 1961). As before the representative consumer's utility function is as follows:

$$U(\boldsymbol{q}, Y, \boldsymbol{p}) = \boldsymbol{q}'\boldsymbol{\delta} - \frac{1}{2}\boldsymbol{q}'M\boldsymbol{q} - \phi(Y - \boldsymbol{q}'\boldsymbol{p}) \tag{3.1}$$

for $\boldsymbol{q}$ the representative consumer's consumption bundle; ϕ the representative consumer's price sensitivity, $\boldsymbol{\delta}$ the $N \times 1$ vector of initial marginal product utilities, M the $N \times N$ positive definite Hessian matrix of the representative consumer's utility function, and $\boldsymbol{p}$ the $N \times 1$ price vector. Y refers to the scalar income/wealth of the representative consumer.

Firms set prices in the second stage of a two-stage profit maximisation game. The optimal price, equilibrium quantity, and equilibrium level of profits are considered in the same two instances as before: where a set of N single-product firms define prices individually, and where a single multi-product monopolist does so for its N goods. Their outcomes appear in (1.25), (1.29), (1.28), and (1.33). For firm-level price-setting, the derivations of the FOCs (unadjusted for corner solutions for simplicity) yield the following:

$$\boldsymbol{p}^* = -\frac{1}{\phi}(\Omega + M^{-1})^{-1}M^{-1}\boldsymbol{\delta} \qquad \boldsymbol{q}(\boldsymbol{p}^*) = \Omega(\Omega + M^{-1})^{-1}M^{-1}\boldsymbol{\delta} \qquad \boldsymbol{\pi}^* = \boldsymbol{p}^* \odot \boldsymbol{q}(\boldsymbol{p}^*) \tag{3.2}$$

for Ω the $N \times N$ matrix representation of the diagonal of M^{-1}. For monopolist price-setting:

$$\boldsymbol{p}^* = -\frac{1}{2\phi}\boldsymbol{\delta} \qquad \boldsymbol{q}(\boldsymbol{p}^*) = \frac{1}{2}M^{-1}\boldsymbol{\delta} \qquad \Pi^* = -\frac{1}{4\phi}\boldsymbol{\delta}'M^{-1}\boldsymbol{\delta} \tag{3.3}$$

In the first stage, firms determine endogenously what product characteristics to ascribe their goods. To understand product design as an optimisation problem for a profit-maximising firm, however, we first need to discuss how product characteristics enter the utility function of the representative consumer.

3.3.2 Characteristics-Based Linear Demand

As in Chapter 2, let initial marginal product utility be linear in product characteristics: $\boldsymbol{\delta} = X\boldsymbol{\beta} + \boldsymbol{v}$, for $X = [\boldsymbol{x}_1 \ldots \boldsymbol{x}_K]$ a N by K matrix whose elements determine how much of a given observable characteristic k good n has; $\boldsymbol{\beta}$ a $K \times 1$ positive vector of the representative consumer's preference weights towards each of K observable characteristics; and $\boldsymbol{v}$ a $N \times 1$ vector of independent, identically distributed, mean zero, unit variance, unobserved errors pertaining to demand shocks and latent product characteristics.[2] Assume $N \geq K$, i.e., there are (weakly) more goods than characteristics.

This is a commonplace functional form assumption for product utility. For product design purposes, it is however insufficiently detailed. The different columns of X may be

[2] "Latent" refers both to variables unobservable to researchers and to variables outside the firm's control.

dependent on each other, making separate optimisation of each unrealistic. For example, we cannot separately choose over a good's mass without addressing how changes in mass impacts their volume; it may not be adequate to consider them as independent choice variables for a profit-maximising firm. If we wish to optimise over each separate characteristic, we must therefore re-state X in such a way that each column isolates a independent lever, while mapping said levers to the original product's characteristics' physical (and/or technological or institutional) rigidities.

Via a QR decomposition[3] of X, we can write $\boldsymbol{\delta} = X\boldsymbol{\beta} + \boldsymbol{v} = ZR\boldsymbol{\beta} + \boldsymbol{v} = Z\tilde{\boldsymbol{\beta}} + \boldsymbol{v}$, for $Z \in \mathbb{R}^{N\times K}$ an orthonormal matrix, $R \in \mathbb{R}^{K\times K}$ an upper-triangular matrix, and $\tilde{\boldsymbol{\beta}} \in \mathbb{R}^{K\times 1}$ an orthogonally-corrected vector of preference weights towards each observable characteristic. Matrix Z rewrites X as orthonormal *attributes*. Matrix R on the other hand is a mechanical mapping; it describes how each attribute relates to each original characteristic.

To get an intuition, let us return to the point above about mass and volume. Mass and volume are related through the concept of density: $density = mass/volume$. Consider mass and volume vectors $\boldsymbol{m}$ and $\boldsymbol{v}$ (the latter not to be confused with the vector of demand shocks and latent product characteristics), for a range of goods of different densities ρ: $X = [\boldsymbol{v}\ \boldsymbol{m}]$:

$$\begin{aligned} X &= ZR \\ &= \begin{bmatrix} \frac{\boldsymbol{v}}{||\boldsymbol{v}||} & \frac{\boldsymbol{m}-\overline{\rho}\boldsymbol{v}}{||\boldsymbol{m}-\overline{\rho}\boldsymbol{v}||} \end{bmatrix} \begin{bmatrix} ||\boldsymbol{v}|| & ||\boldsymbol{v}||\overline{\rho} \\ 0 & ||\boldsymbol{v}||\sigma_\rho \end{bmatrix} \end{aligned} \tag{3.4}$$

for $\overline{\rho} = \sum_{i=1}^{N} w_i\rho_i$ the weighted mean density across all goods, $\sigma_\rho^2 = \sum_{i=1}^{n} w_i(\rho_i - \overline{\rho})^2$ the weighted density variance, and $w_i = v_i^2/||\boldsymbol{v}||^2, \forall i$ the weights.[4] In this way, Z separately isolates two attributes for firms to decide on - product volume and deviation from

[3]For this and other matrix algebra concepts used in this Chapter, see, e.g. Golub & van Loan (1983).
[4]Volume $\boldsymbol{v}$ is squared in the weights because it enters X directly and through $\boldsymbol{m} = \boldsymbol{\rho} \circ \boldsymbol{v}$.

(weighted) mean density $\overline{\boldsymbol{m}} = \overline{\rho}\boldsymbol{v}$, while R codifies how mass of a given good depends on volume plus its own density heterogeneity.

There are a few different methods to orthonormalise vectors. One such method used above, Gram-Schmidt orthogonalisation, retains interpretability: attributes match their respective characteristics *net* of the component explained by the preceding attribute(s). In our example, the first orthonormal direction isolates volume, while the second isolates mass net of its volume-proportional component, that is, net of the mass that would arise if all goods shared the common weighted-average density $\overline{\rho}$.

By restating X such that each characteristic can be independently optimised through underlying attributes, we have made our linear demand model partially characteristics-based. However, endogenous product design not only affects the utility yielded by said products, but also the extent to which product compete. Products may be e.g. closer substitutes the more similar they are in their characteristics. From a utility function perspective, product similarity should play a role in determining how consumption of a given good satiates the consumer's need for said good and those sharing similar attributes. In linear demand, both satiation and substitution effects are driven by the latent Hessian of the utility function $U(\boldsymbol{q})$; in Chapters 1 and 2, said Hessian is represented by M.

What would M look like if we assumed it to be also partially a function of observable product characteristics? Let $\hat{M} = E(M|X)$ and note the following:

Lemma 3.1*: Any Hessian matrix that depends in part on observed product characteristics X and imposes no arbitrary structure on unobserved dimensions admits the following representation:*

$$\hat{M} = ZDZ' + \rho(I - ZZ') \tag{3.5}$$

for D the $K \times K$ reduced Hessian in characteristic coordinates, and ρ a scalar governing the portion of M unexplained by Z.[5]

The matrix D summarises how the observed characteristics contribute to the slope of demand and interact with one another. The scalar ρ effectively weighs the relevance of a baseline-differentiation term.

We can go a little further. The $\hat{M}$ can be re-written as $\hat{M} = \rho I + Z\tilde{D}Z'$, for $\tilde{D} = D - \rho I$. As a symmetric matrix, we can write $\tilde{D} = U\Gamma U'$ via spectral decomposition. At last, we may write:

$$\hat{M} = ZDZ' + \rho(I - ZZ') = (ZU)\Gamma(ZU)' + \rho I = S\Gamma S' + \rho I = \sum_{k=1}^{K} \gamma_k \boldsymbol{s}_k \boldsymbol{s}_k' + \rho I \quad (3.6)$$

for U the orthogonal matrix whose columns are the eigenvectors of $\tilde{D}$, $\Gamma = \text{diag}(\gamma_1, \ldots, \gamma_K)$ the diagonal matrix of its eigenvalues, and $S = ZU$ the resulting orthonormal matrix of attribute directions in product space.

It is trivial to see that matrix $S = ZU$ is also an orthonormal matrix. In fact, matrix S is just another orthonormalised representation for X: $X = ZR = SU^{-1}R$. Matrix S is a re-statement of Z in a new orthonormal basis of the same span - we therefore retain the *attribute* naming. However, we must now keep in mind that an attribute is in fact a linear combination of different characteristics, and some attributes may load the same characteristics in opposite ways. There is no simple way to capture this.

Meanwhile, Γ is a diagonal matrix whose elements pertain to the *competitive salience* of each attribute; these are measures of the relative contribution of differentiation in a given attribute $\boldsymbol{s}_k$ to product positioning in the market, in the style of Bordalo, Gennaioli,

[5] Proofs of all Lemmas may be found in the Appendix.

and Shleifer (2013).

The matrix $\hat{M}$ is thus divided between a baseline idiosyncratic differentiation term ρI and a term measuring competition explainable from observable attributes firms can control. Note that M is positive definite if and only if $\rho + \gamma_K > 0$. I assume $\gamma_k > 0, \ \forall k \in \mathcal{K}$ to be the case throughout this Chapter - positive definiteness will then be automatic if at least one column exists. In general, $\gamma_p \neq \gamma_q, \ \forall p \neq q = 1, \ldots, K$. For simplicity, let $\rho = 1$. To simplify notation, going forward we drop the hat in $\hat{M}$.

This way of formatting M has multiple remarkable features that make it useful both to study the theoretical implications of assuming the relevance of product characteristics for competition, but also for empirical estimation. More on empirical estimation in the Appendix. Additional details on (i) how to obtain S if X is unobserved but we know M and (ii) how $\hat{M}$ can be tied to $\mathbb{E}(\boldsymbol{\delta}\boldsymbol{\delta}')$ - i.e. the second raw moment of $\boldsymbol{\delta}$ - can also be found in the Appendix.

Before we continue, note:

Lemma 3.2: *Product utility $\boldsymbol{\delta} = X\beta + \boldsymbol{v} = S\boldsymbol{b} + \boldsymbol{v}$, for $\boldsymbol{b} = U^{-1}R\beta > 0$ the $K \times 1$ vector of attribute utilities.*

I incorporate also a quadratic cost per product characteristic within a good: $c_n(X) = \frac{1}{2}||\boldsymbol{x}_n||^2$. It can be shown that $c_n(\boldsymbol{x}_n)$ can be written in terms of S:

$$c_n(\boldsymbol{x}_n) = \frac{1}{2}||\boldsymbol{x}_n||^2 = \frac{1}{2}\boldsymbol{x}_n'\boldsymbol{x}_n = \frac{1}{2}\boldsymbol{s}_n'(U^{-1}RR'U)\boldsymbol{s}_n = \frac{1}{2}\boldsymbol{s}_n'C\boldsymbol{s}_n, \ \forall n \tag{3.7}$$

for $C = U^{-1}RR'U$ a common term across goods. At last, we can state regarding expected profits $\mathbb{E}\pi$:

Proposition 3.1: *For a full column rank X:*

$$\max_{\{\boldsymbol{x}_i\}_{i\in\mathcal{G}_n}} \sum_{i\in\mathcal{G}_n} \mathbb{E}\pi_i(\boldsymbol{p}^*(X), X) = \max_{\{\boldsymbol{s}_i\}_{i\in\mathcal{G}_n}} \sum_{i\in\mathcal{G}_n} \mathbb{E}\pi_i(\boldsymbol{p}^*(S), S) \tag{3.8}$$

Proof: This result follows from $X = SU^{-1}R$, with $U^{-1}R$ kept fixed over the optimisation. U is full rank by construction, as it reflects the eigenvectors of a positive definite matrix D, which itself follows from a positive definite M. The mapping from X to S is unique if and only if $U^{-1}R$ is full rank, which it will be if R is full rank. R is full rank if and only if X is full column rank. □

Full column rank is an undemanding assumption, requiring only that the observed characteristics not be exactly collinear across products. Put differently, it fails only when at least one characteristic adds no independent variation beyond the others.[6] Therefore, optimising product design via choice of characteristics X can, under reasonable conditions, be said to be equivalent to doing so over orthonormalised attributes S.

Going forward, keep in mind the following definitions. Any attribute vector can be split into two, separately optimisable concepts: *attribute intensity* and *attribute orientation*. Attribute intensity pertains to the length of the attribute vector, e.g. $||\boldsymbol{s}_k||$; attribute orientation is the direction of the unit-normalised attribute vector: $||\boldsymbol{s}_k||^{-1}\boldsymbol{s}_k$. Furthermore, product differentiation will be defined in terms of *observed* characteristics. Horizontal differentiation will be noted where, for e.g. a product pair, one of the goods has a relatively greater amount of a given attribute, and the other a relatively greater amount in another attribute. This can be best related to differences in attribute orientation. Vertical differentiation on the other hand will be noted where, for the same product pair, one good has (weakly) more of all attributes than the other. This is most observable in differences

[6] Note that for X to be full column rank, it is necessary (but not sufficient) that $K \leq N$, as assumed above.

in attribute intensity. In both orderings, differentiation is qualitatively the same regardless of whether we define it in terms of attributes $\boldsymbol{s}$ or characteristics $\boldsymbol{x}$. The model otherwise assumes horizontal differentiation in unobserved characteristics, meaning that a good with weakly less of all attributes relative to all other goods can still be observed obtaining residual demand.

3.4 Optimal Product Design under Monopoly Pricing

3.4.1 One-Attribute Case

The simplest case is that of a multi-product monopolist optimising over a single attribute ($K = 1$) - we will use it to build intuition. We will prove the following:

Proposition 3.2: *A multi-product product-designing monopolist optimising its products given a single attribute is indifferent in the extent of vertical differentiation across its goods, but never designs horizontal differentiation. Total attribute intensity across goods is increasing in attribute utility and decreasing in price sensitivity, the attribute cost coefficient, and attribute competitive salience.*

Proof: For a single attribute, the Sherman-Morrison–Woodbury formula (Sherman and Morrison, 1949; Woodbury, 1950) simplifies:

$$M^{-1} = I_N - \frac{\gamma \boldsymbol{s}\boldsymbol{s}'}{1 + \gamma ||\boldsymbol{s}||^2} \tag{3.9}$$

Define $\boldsymbol{\delta} = \beta \boldsymbol{x} = (\beta ||\boldsymbol{x}||)||\boldsymbol{x}||^{-1}\boldsymbol{x} = (\beta r)\ \boldsymbol{z} = b\boldsymbol{s}$. In this case, $\boldsymbol{s} = |\boldsymbol{z}|$ as $\boldsymbol{z}\tilde{\beta} = \boldsymbol{z}ub \Leftrightarrow b = \frac{1}{u}\tilde{\beta}$; $||u|| = |u| = 1$ necessarily as it must be orthonormal. Both $\boldsymbol{u} = \pm 1$ and $\boldsymbol{r} = ||\boldsymbol{x}||$ are held fixed.

Note that:

Lemma 3.3: *Expected monopoly profit* $\mathbb{E}\Pi$ *is independent of the idiossyncratic error* $\boldsymbol{v}$.

We therefore define $\boldsymbol{\delta} = b\boldsymbol{s}$ going forward WLOG. Then:

$$\begin{aligned}
\mathbb{E}\Pi &= (-\frac{1}{4\phi})\boldsymbol{\delta}' M^{-1}\boldsymbol{\delta} - \frac{1}{2}||\boldsymbol{x}||^2 \\
&= (-\frac{1}{4\phi})b^2\boldsymbol{s}'(I_N - \frac{\gamma \boldsymbol{s}\boldsymbol{s}'}{1+\gamma||\boldsymbol{s}||^2})\boldsymbol{s} - \frac{1}{2}c||\boldsymbol{s}||^2 \\
&= (-\frac{1}{4\phi})\frac{b^2||\boldsymbol{s}||^2}{1+\gamma||\boldsymbol{s}||^2} - \frac{1}{2}c||\boldsymbol{s}||^2
\end{aligned} \tag{3.10}$$

Notice how, in this context, $\mathbb{E}\Pi$ is shown to be dependent solely on the attribute intensity of $\boldsymbol{s}$. In other words, for a multi-product monopolist whose goods have a single characteristic, the distribution of said characteristic across the goods is not important. It is equally aggregate-profit-maximising for the monopolist to sell a single all-encompassing good n, a set of goods vertically differentiated in the set of attributes firms can control, or a range of fully-homogenous goods within the same set of attributes.[7]

It can be shown that:

$$||\boldsymbol{s}||^* = \arg\max_{||\boldsymbol{s}||} \mathbb{E}\boldsymbol{\Pi}(\boldsymbol{p}^*, \boldsymbol{s}) = \begin{cases} \frac{1}{\sqrt{\gamma}}\sqrt{-\frac{b^2}{2c\phi} - 1} & \text{if } c < -b^2/2\phi \\ 0 & \text{otherwise} \end{cases} \tag{3.11}$$

We can then work backwards to see what this implies for X. For a single attribute, $\boldsymbol{x} = (\frac{\boldsymbol{x}}{||\boldsymbol{x}||})(||\boldsymbol{x}||) = \boldsymbol{z}r$. For S to remain orthonormal, $|u| = 1$. Therefore, $b = U^{-1}R\beta = r\beta$. Thus, we can find out what $\boldsymbol{x}^*$ looks like from the equality: $\beta\boldsymbol{x}^* = b\boldsymbol{s}^* \Leftrightarrow \boldsymbol{x}^* = r\boldsymbol{s}^*$. □

---

[7] Equivalently and respectively: $\boldsymbol{s}^* = ||\boldsymbol{s}||^*\boldsymbol{e}_n$; e.g., $\boldsymbol{s}^* = ||\boldsymbol{s}||^*\sqrt{6/(N(N+1)(2N+1))} \cdot [N\ N-1\ \dots\ 1]'$); $\boldsymbol{s}^* = ||\boldsymbol{s}||^*/\sqrt{N} \cdot \boldsymbol{1}$.

3.4.2 Attribute Exclusivity

In this section, I will prove the following:

Proposition 3.3: *A multi-product monopoly optimising the design of its products given multiple exclusive attributes will load on these attributes in a manner proportional to their cost-weighted attribute utility. Attribute intensity across goods will be increasing in attribute utility and decreasing in price sensitivity, the attribute cost coefficient, and attribute competitive salience.*

Exclusivity means each attribute K is set by (and loads on) exactly one good n:

$$\boldsymbol{s}_k = r_{nk}\boldsymbol{e}_k \qquad \Rightarrow \qquad \boldsymbol{s}_n = \boldsymbol{r} \cdot (\sum_{k \in \mathcal{A}_n} \boldsymbol{e}_k) = \boldsymbol{r}_n \tag{3.12}$$

Define $\mathcal{A}_n$ as the set of all attributes exclusive to n. This allows the following simplification through the Sherman-Morrison–Woodbury formula:

$$M = I_N + \text{diag}(\sum_{k \in \mathcal{A}_1} \gamma_k r_{1k}^2, \ldots, \sum_{k \in \mathcal{A}_N} \gamma_k r_{Nk}^2) \tag{3.13}$$

By construction, we have $\delta_n = \sum_{k \in \mathcal{A}_n} b_k r_{nk},\ \forall n$. Plugging the diagonal M into the demand function and then calculating optimal prices gives, for each firm n:

$$\begin{aligned} q_n^* &= \frac{\delta_n}{2(1 + \sum_{k \in \mathcal{A}_n} \gamma_k r_{nk}^2)} \\ p_n^* &= -\frac{1}{2\phi}\delta = -\frac{1 + \sum_{k \in \mathcal{A}_n} \gamma_k r_{nk}^2}{\phi} q_n^* \end{aligned} \tag{3.14}$$

Product design over exclusive attributes separates each good into a separate, independent corner of the market. Hence the produce design problem decouples by good:

$$\begin{aligned}
&\max_{\{r_{nk}\}_{n,k}} \mathbb{E}\Pi \\
&= \max_{\{r_{nk}\}_{k\in\mathcal{A}_n}} \sum_{n=1}^{N} \mathbb{E}\pi_n(\boldsymbol{r}_n) \\
&= \sum_{n=1}^{N} \max_{\{r_{nk}\}_{k\in\mathcal{A}_n}} -\frac{1}{\phi}(1+\sum_{k\in\mathcal{A}_n}\gamma_k r_{nk}^2)\left(\frac{\sum_{k\in\mathcal{A}_n} b_k r_{nk}}{2(1+\sum_{k\in\mathcal{A}_n}\gamma_k r_{nk}^2)}\right)^2 - \frac{1}{2}\boldsymbol{r_n'}C\boldsymbol{r_n}
\end{aligned} \tag{3.15}$$

As before, let $\hat{r}_k = \sqrt{\gamma_k} r_k$, $\hat{b}_k = \frac{1}{\sqrt{\gamma}} b_k$, and $\hat{C} = \Gamma^{-1/2} C \Gamma^{-1/2}$. Then, let $\hat{\boldsymbol{r}}_n = t_n \boldsymbol{d}_n$, for $t_n = ||\hat{\boldsymbol{r}}_n|| = ||\boldsymbol{\gamma}^{\circ 1/2}\boldsymbol{r}_n||$ and $||\boldsymbol{d}_n|| = 1$, these referring only to the space where $k \in \mathcal{A}_n$, $\forall n$. We obtain:

$$\mathbb{E}\pi_n = -\frac{1}{4\phi}\frac{(\hat{\boldsymbol{b}}'\hat{\boldsymbol{r}}_n)^2}{(1+\hat{\boldsymbol{r}}_n'\hat{\boldsymbol{r}}_n)} - \frac{1}{2}\hat{\boldsymbol{r}}_{\boldsymbol{n}}'\hat{C}\hat{\boldsymbol{r}}_{\boldsymbol{n}} = -\frac{1}{4\phi}\frac{t_n^2(\hat{\boldsymbol{b}}'\boldsymbol{d}_n)^2}{(1+t_n^2)} - \frac{1}{2}t_n^2\boldsymbol{d}_n' C\boldsymbol{d}_n \tag{3.16}$$

Let us write $\hat{\boldsymbol{b}}_n = \sum_{k\in\mathcal{A}_n}\hat{\boldsymbol{b}}\boldsymbol{e}_k$ WLOG: $\hat{\boldsymbol{b}}'\hat{\boldsymbol{r}}_n = \hat{\boldsymbol{b}}_n'\hat{\boldsymbol{r}}_n$. Note then how, via the Cauchy-Schwarz inequality, $(\hat{\boldsymbol{b}}_{\boldsymbol{n}}'\boldsymbol{d}_{\boldsymbol{n}})^2 \leq (\hat{\boldsymbol{b}}_{\boldsymbol{n}}' C^{-1}\hat{\boldsymbol{b}}_{\boldsymbol{n}})(\boldsymbol{d}_{\boldsymbol{n}}' C\boldsymbol{d}_{\boldsymbol{n}})$, holding with equality if $\boldsymbol{d}_n || \{C^{-1}\hat{\boldsymbol{b}}_n\}_{k\in\mathcal{A}_n}$. It thus follows that, in rehashing the same approach as in the previous section, the optimal $\boldsymbol{d}_n$ is as follows:

$$\boldsymbol{s}_n^* = \boldsymbol{r}_n^* = t_n \cdot \Gamma^{-1/2}\boldsymbol{d}_n^* = t_n \cdot \Gamma^{-1/2}\frac{\{\hat{C}^{-1}\hat{\boldsymbol{b}}_n\}_{k\in\mathcal{A}_n}}{||\{\hat{C}^{-1}\hat{\boldsymbol{b}}_n\}_{k\in\mathcal{A}_n}||} = t_n \cdot \frac{\{C^{-1}\boldsymbol{b}_n\}_{k\in\mathcal{A}_n}}{||\{\Gamma^{1/2}C^{-1}\boldsymbol{b}_n\}_{k\in\mathcal{A}_n}||} \tag{3.17}$$

What about t_n? As before, we can plug in $\boldsymbol{d}_n^*$ and perform the derivative of the resulting function in t^2, obtaining:

$$-\frac{(1/4\phi)\boldsymbol{b}_n' C^{-1}\boldsymbol{b}_n}{(1+t_n^2)^2} = \frac{1}{2} \Leftrightarrow t_n^{2*} = \max\left\{\sqrt{-\frac{1}{2\phi}\boldsymbol{b}_n' C^{-1}\boldsymbol{b}_n} - 1, 0\right\} \tag{3.18}$$

Note that $t^* \geq 0$ by construction, but $\boldsymbol{s}_n$ need not be element-by-element positive. This concludes our proof. □

3.4.3 Attribute Non-Exclusivity

Let us now consider the multi-attribute case under non-exclusivity. We will prove the following:

Proposition 3.4: *A multi-product product-designing monopolist optimising its products given multiple attributes does not pursue horizontal differentiation. It is indifferent to the number of goods it produces or the extent of their vertical differentiation; attribute intensity across goods is proportional to their cost-weighted attribute utility; increasing in attribute utility and decreasing in price sensitivity, the attribute cost coefficient, and attribute competitive salience.*

Proof: Consider the pair of attribute vectors $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, describing how much of attributes i and j each good has. By the Cauchy-Schwarz inequality and orthonormality of the initial representation, $\boldsymbol{s}_i'\boldsymbol{s}_j \in [-1, 1]$, with equality only attainable if $\boldsymbol{s}_i || \boldsymbol{s}_j$. To analyse how a change in $\boldsymbol{s}_i'\boldsymbol{s}_j$, $\forall i, j$ affects expected profits in our model without affecting any other attribute pair (including those which relate to either of these attributes), we consider shifts that allow $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ to remain orthogonal to all other pairs. We can do this by varying the pair along the 2-D plane that is orthogonal to the span of the remaining vectors, setting, e.g.:

$$\boldsymbol{s}_i = \boldsymbol{e}_i \qquad \boldsymbol{s}_j = \cos\theta \boldsymbol{e}_j + \sin\theta \boldsymbol{e}_i, \qquad \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}] \tag{3.19}$$

This can be done WLOG. We can organise any pair of $\boldsymbol{s}$ in this form as (i) their multiplication remains the same - we have only rearranged what part of $\boldsymbol{s}_i'\boldsymbol{s}_j$ stays on which side, in the same way that e.g. $12 = 2 \times 6 = 3 \times 4$ - and (ii) it does not affect any other attribute pair.

The first point can be generalised by noting that any $\boldsymbol{s}_i'\boldsymbol{s}_j$ can be re-written as a function of the maximum value it may take (1 in this case since we are keeping attribute intensity across all columns of S fixed) and a parameter θ:

$$\boldsymbol{s}_i'\boldsymbol{s}_j = \sin\theta \in [-1, 1] \tag{3.20}$$

The split into $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, with $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ follows.

To see the second point: for any $k \neq i, j$, assume we start with $\boldsymbol{s}_k \perp \boldsymbol{e}_i$ and $\boldsymbol{s}_k \perp \boldsymbol{e}_j$, such that all $\boldsymbol{s}$ are representationally orthogonal to each other prior to optimisation. Since $\boldsymbol{s}_i(\theta)$ and $\boldsymbol{s}_j(\theta)$ lies in the plane spanned by $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$, we have, $\forall i, j \neq k = 1, \ldots, K$ and θ:

$$\boldsymbol{s}_k'\boldsymbol{s}_i(\theta) = \boldsymbol{s}_k'\boldsymbol{e}_i = 0 \qquad \boldsymbol{s}_k'\boldsymbol{s}_j(\theta) = \cos\theta \boldsymbol{s}_k'\boldsymbol{e}_j + \sin\theta \boldsymbol{s}_k'\boldsymbol{e}_i = 0 \tag{3.21}$$

The proof is trivial: if $\boldsymbol{s}_k'\boldsymbol{e}_i = 0$, true by construction, then (i) $\boldsymbol{s}_k'\boldsymbol{s}_i(\theta) = 0$; (ii) $\sin\theta\boldsymbol{s}_k'\boldsymbol{e}_i = 0$; and (iii) for $\boldsymbol{s}_k'\boldsymbol{s}_j(\theta)$ to be 0 for one θ (such that orthogonality holds at some representation of S), then $\cos\theta\boldsymbol{s}_k'\boldsymbol{e}_j = 0$, which must be true for any θ.

Consider the Sherman–Morrison–Woodbury formula for matrices:

$$M^{-1} = (I_N + \sum_{k=1}^{K} \gamma_k \boldsymbol{s}_k \boldsymbol{s}_k')^{-1} = (I + S\Gamma S')^{-1} = I - S(\Gamma^{-1} + S'S)^{-1}S' \tag{3.22}$$

In respect of partial differentiation, we can thus allow $\boldsymbol{s}_i'\boldsymbol{s}_j$ to vary, while every other pair stays at zero. This means that $\Gamma^{-1} + S'S$ remains block diagonal in what is relevant for optimisation. If we place i and j as the first columns S:

$$\Gamma^{-1} + S'S = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}, \qquad A = \begin{bmatrix} \frac{1}{\gamma_i} + 1 & \boldsymbol{s}_i'\boldsymbol{s}_j \\ \boldsymbol{s}_i'\boldsymbol{s}_j & \frac{1}{\gamma_j} + 1 \end{bmatrix} \qquad B = \begin{bmatrix} \frac{1}{\gamma_3} + 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\gamma_N} + 1 \end{bmatrix} \tag{3.23}$$

This means:

$$(\Gamma^{-1} + S'S)^{-1} = \begin{bmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{bmatrix} \qquad \text{for} \qquad A^{-1} = \frac{1}{\Delta}\begin{bmatrix} \frac{1}{\gamma_j}+1 & -\boldsymbol{s}_i'\boldsymbol{s}_j \\ -\boldsymbol{s}_i'\boldsymbol{s}_j & \frac{1}{\gamma_i}+1 \end{bmatrix} \tag{3.24}$$

for $\Delta = (1/\gamma_i + 1)(1/\gamma_j + 1) - (\boldsymbol{s}_i'\boldsymbol{s}_j)^2$. Since we have that $\boldsymbol{s}_i'\boldsymbol{s}_j \in [-1, 1]$, Δ is strictly positive. From here, note that, from (3.3):

$$\begin{aligned} \mathbb{E}\Pi &= (-\frac{1}{4\phi})\boldsymbol{\delta}' M^{-1}\boldsymbol{\delta} - \frac{1}{2}\sum_{n=1}^{N} ||\boldsymbol{x}_n||^2 \\ &= (-\frac{1}{4\phi})\Big[||\boldsymbol{\delta}||^2 - \boldsymbol{\delta}' S(\Gamma^{-1} + S'S)^{-1} S'\boldsymbol{\delta}\Big] - \frac{1}{2}\sum_{n=1}^{N} \boldsymbol{s_n}' C \boldsymbol{s_n} \\ &= (-\frac{1}{4\phi})\Big[\boldsymbol{b}' S' S \boldsymbol{b} - \boldsymbol{b}' S' S(\Gamma^{-1} + S'S)^{-1} S' S \boldsymbol{b}\Big] - \frac{1}{2}\sum_{n=1}^{N} \boldsymbol{s_n}' C \boldsymbol{s_n} \end{aligned} \tag{3.25}$$

reflecting the fact that, with the exception of the cost term, $\mathbb{E}\boldsymbol{\Pi}$ depends only on attribute intensities ($||\boldsymbol{s}_i||$, $\forall i = 1, \ldots K$) and angles (how attribute orientation differs relative to others, rather than in absolute: $\boldsymbol{s}_i'\boldsymbol{s}_j$, $\forall i, j = 1, \ldots K$).

Regarding the cost term, we can express it in terms of attribute vectors (rather than per-good vectors) as follows:

$$\frac{1}{2}\sum_{n=1}^{N} \boldsymbol{s}_n' C_n \boldsymbol{s}_n = \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{l=1}^{K} \{C_n\}_{kl} s_{nk} \boldsymbol{s}_n l = \frac{1}{2}\sum_{k=1}^{K}\sum_{l=1}^{K} \boldsymbol{s}_k' D_{kl} \boldsymbol{s}_l \tag{3.26}$$

with $D_{kl}, \forall k, l = 1, \ldots K$ a diagonal matrix which places the $C_{n,kl}$ element in the n-th diagonal. The expression can be further simplified for a symmetric and identical C across all goods. Partialling out all terms in the cost term which depend on attributes i and j and accounting for orthogonality, we get:

$$\left\{\sum_{n=1}^{N} c_n(X)\right\}_{i,j} = \frac{1}{2} C_{ii} ||\boldsymbol{s}_i||^2 + \frac{1}{2} C_{jj} ||\boldsymbol{s}_j||^2 + C_{ij} \boldsymbol{s}_i' \boldsymbol{s}_j \tag{3.27}$$

We can then partial out the attribute vectors' contribution to $\mathbb{E}\Pi$:

$$\mathbb{E}\Pi_{\{i,j\}} = \frac{b_i^2 \frac{1}{\gamma_i}\left[\frac{1}{\gamma_j} + 1 - (\boldsymbol{s}_i'\boldsymbol{s}_j)^2\right] + b_j^2 \frac{1}{\gamma_j}\left[\frac{1}{\gamma_i} + 1 - (\boldsymbol{s}_i'\boldsymbol{s}_j)^2\right] + 2b_i b_j \frac{1}{\gamma_i}\frac{1}{\gamma_j}\boldsymbol{s}_i'\boldsymbol{s}_j}{-4\phi\Delta} - \left\{\sum_{n=1}^{N} c_n(X)\right\}_{i,j} \tag{3.28}$$

Differentiating this expression relative to θ:

$$\frac{\partial \mathbb{E}\Pi_{\{i,j\}}}{\partial\theta} = \frac{\left[2b_i b_j \frac{1}{\gamma_i}\frac{1}{\gamma_j}\left[\left(\frac{1}{\gamma_i}+1\right)\left(\frac{1}{\gamma_j}+1\right) + \sin^2\theta\right] - 2\sin\theta\left[b_i^2\left(\frac{1}{\gamma_i}\right)^2\left(\frac{1}{\gamma_j}+1\right) + b_j^2\left(\frac{1}{\gamma_j}\right)^2\left(\frac{1}{\gamma_i}+1\right)\right]\right]\cos\theta}{-4\phi\Delta^2} - C_{ij}\cos\theta \tag{3.29}$$

By definition, $\cos\theta \geq 0$. Therefore, we observe stationary points at $\cos\theta = 0$, equivalent to the $\boldsymbol{s}_i'\boldsymbol{s}_j = \pm 1$ cases. In other words, full or exact opposite orientation alignment are candidate optima. However, note how, for a sufficiently small C, $\mathbb{E}\Pi_{\{i,j\}}$ is greater at every positive value of θ than at the corresponding negative value; $\frac{\partial \mathbb{E}\Pi_{\{i,j\}}}{\partial\theta}$ is also greater at every negative θ than at every positive; and at $\theta = 0$, $\frac{\partial \mathbb{E}\Pi_{\{i,j\}}}{\partial\theta} > 0$. For a sufficiently small C, this rules out global optima in the region where $\boldsymbol{s}_i'\boldsymbol{s}_j < 0$.

This result by itself does not rule out $\boldsymbol{s}_i'\boldsymbol{s}_j < 0$ or an interior global solution, i.e. some degree of horizontal differentiation. We explore this further in Section 3.6., when we introduce asymmetric costs. There, we prove, from a different perspective, that for the multi-product monopolist $\boldsymbol{s}_1||\ldots||\boldsymbol{s}_N$ with positive proportionality. If this is true, then there are no interior solutions in the attribute vectors: $\boldsymbol{s}_i'\boldsymbol{s}_j = \pm 1$ at most.

Assume this to be the case here.

This leaves two remaining questions: (i) what is the optimal orientation that all columns of S should be parallel to; and (ii) what is the optimal attribute intensity. Regarding the first of these, based on the parallelism of the $\boldsymbol{s}$ vectors, allow every $\boldsymbol{s}_k = r_k\boldsymbol{y}$, for

r_k a scalar and $\boldsymbol{y} > 0$ the common factor, $||\boldsymbol{y}|| = 1$. This setup effectively reduces the multi-factor case:

$$\boldsymbol{\delta} = \sum_{k=1}^{K} b_k s_k = \sum_{k=1}^{K} b_k r_k \boldsymbol{y} = (\boldsymbol{b}'\boldsymbol{r})\boldsymbol{y} \tag{3.30}$$

$$M^{-1} = I - \frac{\sum_{k=1}^{K}(\gamma_k r_k)^2}{1+\sum_{k=1}^{K}(\gamma_k r_k^2)}\boldsymbol{y}\boldsymbol{y}' \tag{3.31}$$

so that:

$$\begin{aligned} \mathbb{E}\Pi &= (-\frac{1}{4\phi})\frac{\sum_{k=1}^{K}(b_k r_k)^2}{1+\sum_{k=1}^{K}(\gamma_k r_k^2)} - \frac{1}{2}\sum_{i=1}^{N}\boldsymbol{s_n'}C\boldsymbol{s_n} \\ &= (-\frac{1}{4\phi})\frac{(\boldsymbol{b}'\boldsymbol{r})^2}{1+\boldsymbol{r}'\Gamma\boldsymbol{r}} - \frac{1}{2}||\boldsymbol{y}||^2\boldsymbol{r}'C\boldsymbol{r} \\ &= (-\frac{1}{4\phi})\frac{(\hat{\boldsymbol{b}}'\hat{\boldsymbol{r}})^2}{1+\hat{\boldsymbol{r}}'\hat{\boldsymbol{r}}} - \frac{1}{2}\hat{\boldsymbol{r}}'\hat{C}\hat{\boldsymbol{r}} \end{aligned} \tag{3.32}$$

for $\hat{r}_k = \sqrt{\gamma_k} r_k$ and $\hat{b}_k = \frac{1}{\sqrt{\gamma_k}} b_k$, $\forall k = 1, \ldots, K$. We also use $\hat{C} = \Gamma^{-1/2}C\Gamma^{-1/2}$, $\forall k$. Thus, we have, via a further decomposition, $\hat{\boldsymbol{r}} = t\boldsymbol{d}$, with $t = ||\hat{\boldsymbol{r}}|| = ||\boldsymbol{\gamma}^{\circ 1/2}\boldsymbol{r}|| > 0$ setting intensity, and $\boldsymbol{d}$ setting sign and orientation, with $||\boldsymbol{d}|| = 1$:

$$\mathbb{E}\Pi = (-\frac{1}{4\phi})\frac{(\hat{\boldsymbol{b}}'\hat{\boldsymbol{r}})^2}{1+\hat{\boldsymbol{r}}'\hat{\boldsymbol{r}}} - \frac{1}{2}\hat{\boldsymbol{r}}'\hat{C}\hat{\boldsymbol{r}} = (-\frac{1}{4\phi})\frac{t^2(\hat{\boldsymbol{b}}'\boldsymbol{d})^2}{1+t^2} - \frac{1}{2}t^2\boldsymbol{d}'\hat{C}\boldsymbol{d} \tag{3.33}$$

The only relevant choice variable is r_k, $\forall k$, the norm of each vector. What is the optimal $\hat{\boldsymbol{r}}$? Via Cauchy-Schwarz, $(\hat{\boldsymbol{b}}'\boldsymbol{d})^2 \leq (\hat{\boldsymbol{b}}'\hat{C}^{-1}\hat{\boldsymbol{b}})(\boldsymbol{d}'\hat{C}\boldsymbol{d})$, holding with equality if $\boldsymbol{d} \parallel \hat{C}^{-1}\hat{\boldsymbol{b}}$. This is key, as we can show that it maximises $\mathbb{E}\Pi$:

$$\begin{aligned} \mathbb{E}\Pi &\leq (-\frac{1}{4\phi})\frac{t^2(\hat{\boldsymbol{b}}'\hat{C}^{-1}\hat{\boldsymbol{b}})(\boldsymbol{d}'\hat{C}\boldsymbol{d})}{1+t^2} - \frac{1}{2}t^2\boldsymbol{d}'\hat{C}\boldsymbol{d} \\ &= \left((-\frac{1}{4\phi})\frac{t^2(\hat{\boldsymbol{b}}'\hat{C}^{-1}\hat{\boldsymbol{b}})}{1+t^2} - \frac{1}{2}t^2\right)\boldsymbol{d}'\hat{C}\boldsymbol{d} \\ &\propto \left((-\frac{1}{4\phi})\frac{t^2(\hat{\boldsymbol{b}}'\hat{C}^{-1}\hat{\boldsymbol{b}})}{1+t^2} - \frac{1}{2}t^2\right)\hat{\boldsymbol{b}}\hat{C}^{-1}\hat{\boldsymbol{b}} \end{aligned} \tag{3.34}$$

which is independent of $\boldsymbol{d}$, such that the expression has a zero gradient in that variable.

In other words, for any fixed t:

$$\boldsymbol{r}^* = t \cdot \Gamma^{-1/2}\boldsymbol{d}^* = t \cdot \Gamma^{-1/2}\frac{\hat{C}^{-1}\hat{\boldsymbol{b}}}{||\hat{C}^{-1}\hat{\boldsymbol{b}}||} = t \cdot \frac{C^{-1}\boldsymbol{b}}{||\Gamma^{1/2}C^{-1}\boldsymbol{b}||} \tag{3.35}$$

C need not be diagonally dominant, hence the sign uncertainty for $\boldsymbol{r}^*$. What about t? Working from the above, we can plug in $\boldsymbol{d}^*$ and perform the derivative of the resulting (concave) function in t^2, obtaining:

$$-\frac{(1/4\phi)\boldsymbol{b}'C^{-1}\boldsymbol{b}}{(1+t^2)^2} = \frac{1}{2} \Leftrightarrow t^{2*} = \max\left\{\sqrt{-\frac{1}{2\phi}\boldsymbol{b}'C^{-1}\boldsymbol{b}} - 1, 0\right\} \tag{3.36}$$

We reach a result equivalent to the one-attribute case - $\mathbb{E}\boldsymbol{\Pi}$ is independent of $\boldsymbol{y}$. We can conclude that, where full or inverse alignment is optimal, for this setting too is it equally aggregate-profit-maximising for the monopolist to sell a single all-encompassing good n, a vertically-differentiated set of goods in the set of observed attributes, or a range of homogenous goods in the set of observable attributes. For as long as $||\boldsymbol{y}|| = 1$, aggregate profit is maximised independently by every $\boldsymbol{r}$. □

3.4.4 Discussion

What is the intuition for our results? Equation (3.25) provides a hint. There are three mechanisms driving effects of product design on expected profits. First, there is an implicit complementarity in how attributes are stacked. The norm of the initial product utility vector is highest when goods are parallel. For e.g. the two-attribute case:

$$||\boldsymbol{\delta}||^2 = \boldsymbol{b}'S'S\boldsymbol{b} = b_1(b_1 + \boldsymbol{s}_i'\boldsymbol{s}_j b_2) + b_2(\boldsymbol{s}_i'\boldsymbol{s}_j b_1 + b_2) \qquad \frac{\partial||\boldsymbol{\delta}||^2}{\partial \boldsymbol{s}_i'\boldsymbol{s}_j} = 2b_1 b_2 > 0 \tag{3.37}$$

This also drives indifference in the number of goods, as gains in utility through one good or the other are equivalent.

The second mechanism is competitive pressure. Cross-price effects are highest at $\boldsymbol{s}_i'\boldsymbol{s}_j = \pm 1$ and attributes are (anti-)parallel (see (3.24)). As can be seen in (3.25), if $\Gamma^{-1} = 0$, the effect would cancel. With $\Gamma^{-1} > 0$, however, this effect never compensates the first mechanism.

Further alignment has a net positive result on optimal demand - this follows almost immediately from a well-behaved (strictly concave) utility function. Alignment stops the monopolist from spreading attractiveness across different directions and instead lets it push the whole product line towards the direction the consumer likes most. The whole line is pulled toward the same attractive part of product space.

The third mechanism is that of cost c_n. In some cases, elements of C can be negative. This happens when the two attributes act as cost-substitutes: producing more of both together is cheaper than one might infer from adding their standalone costs only, because some of the underlying characteristic adjustments overlap or undo each other. When it happens, this incentivises further alignment.

3.5 Optimal Product Design under Single-Product Firm Pricing

Now consider N single-product firms, one good per firm, each with up to K attributes. Each firm sets its own price. As before, we work with expected profits and define $\boldsymbol{\delta} = S\boldsymbol{b}$ for simplicity. This is WLOG for the optimal attribute orientation proof below, and qualitatively unimportant for our discussion of optimal attribute intensity.

3.5.1 Attribute Exclusivity

We start with multi-attribute product design, but with each attribute being exclusively assigned to only one product. This can be akin to assuming heterogeneity in the product characteristics each firm is allowed control over. From our earlier example with mass and volume, the first firm could be given control over only the volume of their good (first attribute), picking over different standardised materials/formulations; whereas the second firm could set how far to deviate from mean density (second attribute), perhaps through a fixed format which they could then vary in formulation or material composition. In other words, we could define attribute exclusivity such that one is only allowed to make the package bigger at average density, while the other can only change density/formulation holding package size roughly fixed.

I will prove the following:

Proposition 3.5: *A single-product firm optimising the design of its product given multiple exclusive attributes will load on these attributes in a manner proportional to their cost-weighted attribute utility. Attribute intensity across goods will be increasing in attribute utility and decreasing in price sensitivity, the attribute cost coefficient, and attribute competitive salience.*

Proof: The proof mirrors that of the multi-product monopolist setting. Exclusivity means each attribute K is set by (and loads on) exactly one firm n:

$$\boldsymbol{s}_k = r_{nk}\boldsymbol{e}_k \qquad \Rightarrow \qquad \boldsymbol{s}_n = \boldsymbol{r} \cdot (\sum_{k \in \mathcal{A}_n} \boldsymbol{e}_k) = \boldsymbol{r}_n \tag{3.38}$$

Define $\mathcal{A}_n$ as the set of all attributes exclusive to n. This allows the following simplification:

$$M = I_N + \text{diag}(\sum_{k\in\mathcal{A}_1} \gamma_k r_{1k}^2, \dots, \sum_{k\in\mathcal{A}_N} \gamma_k r_{Nk}^2) \tag{3.39}$$

By construction, we have $\delta_n = \sum_{k\in\mathcal{A}_n} b_k r_{nk},\ \forall n$. Plugging the diagonal M into the demand function and then calculating optimal prices gives, for each firm n:

$$\begin{aligned} q_n^* &= \frac{\delta_n}{2(1+\sum_{k\in\mathcal{A}_n}\gamma_k r_{nk}^2)} \\ p_n^* &= -\frac{1}{2\phi}\delta = -\frac{1+\sum_{k\in\mathcal{A}_n}\gamma_k r_{nk}^2}{\phi} q_n^* \end{aligned} \tag{3.40}$$

This is effectively the monopoly setting. Product design over exclusive attributes separates each good into a separate, independent corner of the market. Hence the produce design problem decouples by firm:

$$\max_{\{r_{nk}\}_{k\in\mathcal{A}_n}} \mathbb{E}\pi_n = \max_{\{r_{nk}\}_{k\in\mathcal{A}_n}} -\frac{1}{\phi}(1+\sum_{k\in\mathcal{A}_n}\gamma_k r_{nk}^2)\left(\frac{\sum_{k\in\mathcal{A}_n} b_k r_{nk}}{2(1+\sum_{k\in\mathcal{A}_n}\gamma_k r_{nk}^2)}\right)^2 - \frac{1}{2}\boldsymbol{r_n'}C\boldsymbol{r_n} \tag{3.41}$$

with quadratic attribute costs defined as follows:

$$\frac{1}{2}\boldsymbol{s}_n' C\boldsymbol{s}_n = \frac{1}{2}\boldsymbol{r_n'}C\boldsymbol{r_n} \tag{3.42}$$

As before, let $\hat{r}_k = \sqrt{\gamma_k} r_k$, $\hat{b}_k = \frac{1}{\sqrt{\gamma}} b_k$, and $\hat{C} = \Gamma^{-1/2}C\Gamma^{-1/2}$. Then, let $\hat{\boldsymbol{r}}_n = t_n \boldsymbol{d}_n$, for $t_n = ||\hat{\boldsymbol{r}}_n|| = ||\boldsymbol{\gamma}^{\circ 1/2}\boldsymbol{r}_n||$ and $||\boldsymbol{d}_n|| = 1$, these referring only to the space where $k\in\mathcal{A}_n,\ \forall n$. We obtain:

$$\mathbb{E}\pi_n = -\frac{1}{4\phi}\frac{(\hat{\boldsymbol{b}}'\hat{\boldsymbol{r}}_n)^2}{(1+\hat{\boldsymbol{r}}_n'\hat{\boldsymbol{r}}_n)} - \frac{1}{2}\hat{\boldsymbol{r}}_{\boldsymbol{n}}'\hat{C}\hat{\boldsymbol{r}}_{\boldsymbol{n}} = -\frac{1}{4\phi}\frac{t_n^2(\hat{\boldsymbol{b}}'\boldsymbol{d}_n)^2}{(1+t_n^2)} - \frac{1}{2}t_n^2\boldsymbol{d}_n' C\boldsymbol{d}_n \tag{3.43}$$

Let us write $\hat{\boldsymbol{b}}_n = \sum_{k\in\mathcal{A}_n}\hat{\boldsymbol{b}}\cdot\boldsymbol{e}_k$ WLOG: $\hat{\boldsymbol{b}}'\hat{\boldsymbol{r}}_n = \hat{\boldsymbol{b}}_n'\hat{\boldsymbol{r}}_n$. Note then how, once again with Cauchy-Schwarz, $(\hat{\boldsymbol{b}}_{\boldsymbol{n}}'\boldsymbol{d_n})^2 \leq (\hat{\boldsymbol{b}}_{\boldsymbol{n}}'C^{-1}\hat{\boldsymbol{b}}_{\boldsymbol{n}})(\boldsymbol{d_n}'C\boldsymbol{d_n})$, holding with equality if $\boldsymbol{d}_n||\{C^{-1}\hat{\boldsymbol{b}}_n\}_{k\in\mathcal{A}_n}$. It thus follows that, in rehashing the same approach as in the previous section, the optimal $\boldsymbol{d}_n$ is as follows:

$$\boldsymbol{s}_n^* = \boldsymbol{r}_n^* = t_n \cdot \Gamma^{-1/2} \boldsymbol{d}_n^* = t_n \cdot \Gamma^{-1/2} \frac{\{\hat{C}^{-1} \hat{\boldsymbol{b}}_n\}_{k \in \mathcal{A}_n}}{||\{\hat{C}^{-1} \hat{\boldsymbol{b}}_n\}_{k \in \mathcal{A}_n}||} = t_n \cdot \frac{\{C^{-1} \boldsymbol{b}_n\}_{k \in \mathcal{A}_n}}{||\{\Gamma^{1/2} C^{-1} \boldsymbol{b}_n\}_{k \in \mathcal{A}_n}||} \quad (3.44)$$

What about t_n? As before, we can plug in $\boldsymbol{d}_n^*$ and perform the derivative of the resulting function in t^2, obtaining:

$$-\frac{(1/4\phi)\boldsymbol{b}_n' C^{-1} \boldsymbol{b}_n}{(1 + t_n^2)^2} = \frac{1}{2} \Leftrightarrow t_n^{2*} = \max \left\{ \sqrt{-\frac{1}{2\phi} \boldsymbol{b}_n' C^{-1} \boldsymbol{b}_n} - 1, 0 \right\} \quad (3.45)$$

Once again, $t^* \geq 0$ by construction, but $\boldsymbol{s}_n$ need not be element-by-element positive. This concludes our proof. □

3.5.2 Attribute Non-Exclusivity

Let us now turn to a more complicated setting: multi-attribute design without exclusivity. With non-exclusivity, any firm may load any attribute K.

We will prove the following:

Proposition 3.6*: Competition between single-product price-setting firms optimising the design of its product given multiple attributes reaches a unique symmetric equilibrium. While their starting point matches that of the same setting in the monopoly pricing case - i.e. an optimal attribute vector proportional to their cost-weighted attribute utility - competing firms tilt their attribute vector across greater attribute dimensions. Competition makes firms more balanced in the attributes they load, particularly by emphasizing the attributes with lesser competitive salience over those with lower weighted costs as the number of firms increases.*

Proof: Many of the simplifications that have been made possible by either product exclusivity or joint ownership are no longer available to us. Instead, we take advantage of the symmetry of the firms. Assume a symmetric candidate equilibrium in which all firms choose the same attribute vector:

$$\boldsymbol{s}_n = \boldsymbol{r} = t \cdot \Gamma^{-1/2}\boldsymbol{d} \in \mathbb{R}^K, \ \forall n \tag{3.46}$$

for $||\boldsymbol{d}|| = 1$ and $t = ||\hat{\boldsymbol{r}}|| \geq 0$ as before. First, I will let this be true while allowing a given firm n to deviate from $\boldsymbol{r}$ by setting $\boldsymbol{s}_n = \bar{t} \cdot \Gamma^{-1/2}\boldsymbol{d}$. Define for this purpose $\boldsymbol{t} = \sum_{i \neq n} \boldsymbol{e}_i + \bar{t}\boldsymbol{e}_n$. The goal will be determining the optimal symmetric t^* via $\frac{\partial \mathbb{E}\pi_n}{\partial \bar{t}}|_{\bar{t}=t} = 0$ Second, I will allow a given firm n to deviate by setting $\overline{\boldsymbol{d}}_n$ and verify the conditions under which $\frac{\partial \mathbb{E}\pi_n}{\partial \bar{\boldsymbol{d}}}|_{\bar{\boldsymbol{d}}=\boldsymbol{d}} = 0$.

Note:

$$M = I + S\Gamma S' = I + \hat{S}\hat{S}' = I + (\boldsymbol{t}\boldsymbol{d}')(\boldsymbol{t}\boldsymbol{d}')' = I + \boldsymbol{t}\boldsymbol{d}'\boldsymbol{d}\boldsymbol{t}' = I + \boldsymbol{t}\boldsymbol{t}' \tag{3.47}$$

We can go further. By the Sherman–Morrison–Woodbury formula for the 1-factor case:

$$M^{-1} = I - \frac{\boldsymbol{t}\boldsymbol{t}'}{1 + \bar{t}^2 + (N-1)t^2} \tag{3.48}$$

with $\Omega_{nn} = \frac{1+(N-1)t^2}{1+\bar{t}^2+(N-1)t^2}$ and $\Omega_{-n,-n} = \frac{1+\bar{t}^2+(N-2)t^2}{1+\bar{t}^2+(N-1)t^2}$. Then, for $\boldsymbol{\delta} = \boldsymbol{t}\boldsymbol{d}'\hat{\boldsymbol{b}} = \hat{\boldsymbol{b}}'\boldsymbol{d}\boldsymbol{t}$ and:

$$M^{-1}\boldsymbol{\delta} = \left(I - \frac{\boldsymbol{t}\boldsymbol{t}'}{1 + \bar{t}^2 + (N-1)t^2}\right)\hat{\boldsymbol{b}}'\boldsymbol{d}\boldsymbol{t} = \hat{\boldsymbol{b}}'\boldsymbol{d}\left(I - \frac{\boldsymbol{t}\boldsymbol{t}'}{1 + \bar{t}^2 + (N-1)t^2}\right)\boldsymbol{t} = \frac{1}{1 + \bar{t}^2 + (N-1)t^2}\hat{\boldsymbol{b}}'\boldsymbol{d}\boldsymbol{t} \tag{3.49}$$

we may re-write the optimal price equation as follows, separately defining $\boldsymbol{p}^* =$

$(p_n^* \quad \boldsymbol{p}_{-n}^*)'$, for $\boldsymbol{p}_{-n}^* = p_{-n}^*\mathbf{1}$, due to symmetry:

$$\begin{aligned}
&\boldsymbol{p}^* = -\frac{1}{\phi}(\Omega + M^{-1})^{-1}M^{-1}\boldsymbol{\delta} \\
\Leftrightarrow &(\Omega + M^{-1})\boldsymbol{p}^* = -\frac{1}{\phi}M^{-1}\boldsymbol{\delta} \\
\Leftrightarrow &\begin{pmatrix} 2(1+(N-1)t^2)p_n^* - (N-1)\bar{t}tp_{-n}^* \\ -\bar{t}tp_n^* + (2+2\bar{t}^2+(N-2)t^2)p_{-n}^* \end{pmatrix} = \begin{pmatrix} -\frac{1}{\phi}\hat{\boldsymbol{b}}'\boldsymbol{d}\bar{t} \\ -\frac{1}{\phi}\hat{\boldsymbol{b}}'\boldsymbol{d}t \end{pmatrix}
\end{aligned} \tag{3.50}$$

Solving the system of equations, optimal prices are:

$$p_n^* = -\frac{1}{\phi}\hat{\boldsymbol{b}}'\boldsymbol{d}\bar{t}(\frac{2+2\bar{t}^2+(2N-3)t^2}{\Delta}) \tag{3.51}$$

$$p_{-n}^* = -\frac{1}{\phi}\hat{\boldsymbol{b}}'\boldsymbol{d}t(\frac{\bar{t}^2+2(1+(N-1)t^2)}{\Delta}) \tag{3.52}$$

for $\Delta = 2(1+(N-1)t^2)(2+2\bar{t}^2+(N-2)t^2) - (N-1)\bar{t}^2t^2$. Optimal quantities follow from $\boldsymbol{q}(\boldsymbol{p}^*) = -\phi\Omega\boldsymbol{p}^*$:

$$q_n(\boldsymbol{p}^*) = -\phi\Omega_{nn}p_n^* \qquad q_{-n}(\boldsymbol{p}^*) = -\phi\Omega_{-n,-n}p_{-n}^* \tag{3.53}$$

and expected firm profit will be:

$$\begin{aligned}
\mathbb{E}\pi_n =& \; p_n^*q_n(\boldsymbol{p}^*) - \frac{1}{2}\hat{\boldsymbol{s}}_n'\hat{C}\hat{\boldsymbol{s}}_n \\
=& -\phi\Omega_{nn}p_n^{*2} - \frac{1}{2}\bar{t}^2\boldsymbol{d}'\hat{C}\boldsymbol{d} \\
=& -\frac{1}{\phi}(\frac{1+(N-1)t^2}{1+\bar{t}^2+(N-1)t^2})(\frac{2+2\bar{t}^2+(2N-3)t^2}{\Delta})^2(\hat{\boldsymbol{b}}'\boldsymbol{d}\bar{t})^2 - \frac{1}{2}\bar{t}^2\boldsymbol{d}'\hat{C}\boldsymbol{d}
\end{aligned} \tag{3.54}$$

Then, taking the FOC of $\mathbb{E}\pi_n$ on $\bar{t}$:

$$\begin{aligned}
\left.\frac{\partial\mathbb{E}\pi_n}{\partial\bar{t}}\right|_{\bar{t}=t} &= -\frac{1}{\phi}(\hat{\boldsymbol{b}}'\boldsymbol{d})^2t\frac{2(1+(N-1)t^2)P_N(t)}{(1+Nt^2)^2(2+(N-1)t^2)^3(2+(2N-1)t^2)} - t\boldsymbol{d}'\hat{C}\boldsymbol{d} \\
&= t(F_N(t)(\hat{\boldsymbol{b}}'\boldsymbol{d})^2 - \boldsymbol{d}'\hat{C}\boldsymbol{d})
\end{aligned} \tag{3.55}$$

for $P_N(t) = 4 + 2(5N-4)t^2 + (N-1)(8N-7)t^4 + (N-1)(2N^2-5N+1)t^6$. It can be shown that, for $N \geq 2$, if $F_N(t) > 0$, so $\frac{\partial F_N(t)}{\partial t} < 0$. Any interior symmetric solution must satisfy $\left.\frac{\partial \mathbb{E}\pi_n}{\partial t}\right|_{\bar{t}=t} = 0$ and $P_N(t) \geq 0$ (as $\boldsymbol{d}'\hat{C}\boldsymbol{d} > 0$[8]). Since on that region we have just stated that the first term on the RHS of (3.55) is strictly decreasing in t, the equation $\left.\frac{\partial \mathbb{E}\pi_n}{\partial t}\right|_{\bar{t}=t} = 0$ must have at most one positive solution. We can furthermore show that, for $t^* > 0$ and evaluating at $t = 0$, we need:

$$\begin{aligned}
&(\hat{\boldsymbol{b}}'\boldsymbol{d})^2 F_N(0) \geq \boldsymbol{d}'\hat{C}\boldsymbol{d} \\
\Leftrightarrow & -\frac{1}{2\phi}(\hat{\boldsymbol{b}}'\boldsymbol{d})^2 \geq \boldsymbol{d}'\hat{C}\boldsymbol{d} \\
\Rightarrow & -\frac{1}{2\phi}(\hat{\boldsymbol{b}}'\hat{C}^{-1}\hat{\boldsymbol{b}})(\boldsymbol{d}'\hat{C}\boldsymbol{d}) \geq \boldsymbol{d}'\hat{C}\boldsymbol{d} \\
\Rightarrow & -\frac{1}{2\phi}\boldsymbol{b}'C^{-1}\boldsymbol{b} \geq 1
\end{aligned} \tag{3.56}$$

a condition aligned with the other setups we have considered. Even without a closed-form solution, we have been able to argue for the existence of a unique symmetric optimal attribute intensity.[9]

We now move on to proving that there is a unique symmetric equilibrium attribute orientation. To achieve this, we stick with $\boldsymbol{s}_{-n} = t \cdot \Gamma^{-1/2}\boldsymbol{d}$, and now set, for a firm n, $\boldsymbol{s} = t \cdot \Gamma^{-1/2}\boldsymbol{g}$. Vector $\boldsymbol{g}$ will have a unique structure:

$$\boldsymbol{g} = \cos(\theta)\boldsymbol{d} + \sin(\theta)\boldsymbol{h}, \quad \boldsymbol{h} \perp \boldsymbol{d}, \quad ||\boldsymbol{h}|| = 1 \tag{3.57}$$

such that the deviation we consider is only in orientation. Unlike a format used earlier in the Chapter to analyse attribute angles, this approach allows θ to stay within the range $[0, \pi]$ (0 if $\boldsymbol{g}'\boldsymbol{d} = 1$, π if $\boldsymbol{g}'\boldsymbol{d} = -1$), facilitating interpretation of derivatives relative to this

[8] This is because C is positive definite, since $C = U'X'XU$ and X is full column rank.

[9] Were we to restore $\boldsymbol{v}$ to $\boldsymbol{\delta}$, we would require: $(\hat{\boldsymbol{b}}'\boldsymbol{d})^2 \geq (3N+1)/(2N+2) \xrightarrow{\infty} 1.5$ so that the first term of (3.55) is decreasing in t; and the condition in place of (3.56) be $-\frac{1}{2\phi}((\hat{\boldsymbol{b}}'\boldsymbol{d})^2 - 1) \geq \boldsymbol{d}'\hat{C}\boldsymbol{d}$.

parameter. We can then write M as follows:

$$M(\theta) = \begin{bmatrix} 1+t^2 & t^2(g(\theta)'\boldsymbol{d})\mathbf{1}'_{N-1} \\ t^2(g(\theta)'\boldsymbol{d})\mathbf{1}_{N-1} & I_{N-1}+t^2\mathbf{1}_{N-1}\mathbf{1}'_{N-1} \end{bmatrix} = \begin{bmatrix} 1+t^2 & t^2\cos(\theta)\mathbf{1}'_{N-1} \\ t^2\cos(\theta)\mathbf{1}_{N-1} & I_{N-1}+t^2\mathbf{1}_{N-1}\mathbf{1}'_{N-1} \end{bmatrix} \quad (3.58)$$

We can then show that, for small deviations of $\boldsymbol{g}$ from $\boldsymbol{d}$, the impact on M is second-order:

$$M(\theta) - M(0) = t^2(g(\theta)'\boldsymbol{d} - 1) \begin{bmatrix} 0 & \mathbf{1}'_{N-1} \\ \mathbf{1}_{N-1} & 0 \end{bmatrix} = t^2(\cos(\theta) - 1) \begin{bmatrix} 0 & \mathbf{1}'_{N-1} \\ \mathbf{1}_{N-1} & 0 \end{bmatrix} \quad (3.59)$$

As $\left.\frac{\partial(M(\theta)-M(0))}{\partial\theta}\right|_{\theta=0} \propto -sin(0) = 0$, the effect of a firm deviating from $\boldsymbol{d}$ on M, and therefore M^{-1}, Ω, and $(\Omega + M^{-1})^{-1}$, is negligible. Consider now $\boldsymbol{\delta}$. Its components are:

$$\delta_n(\theta) = t\hat{\boldsymbol{b}}\boldsymbol{g}(\theta) \quad \delta_{-n} = t\hat{\boldsymbol{b}}\boldsymbol{d} \quad (3.60)$$

The effect of a changing attribute orientation on $\boldsymbol{\delta}$ is first-order:

$$\left.\frac{\partial \boldsymbol{\delta}}{\partial\theta}\right|_{\theta=0} = t(\hat{\boldsymbol{b}}'h)\boldsymbol{e}_n \quad (3.61)$$

Bringing these results together:

$$\left.\frac{\partial M^{-1}(\theta)\boldsymbol{\delta}(\theta)}{\partial\theta}\right|_{\theta=0} = t(\hat{\boldsymbol{b}}'h)(\boldsymbol{e}_n - \frac{t^2}{1+Nt^2}\mathbf{1}) \quad (3.62)$$

The second term in brackets is recognisable from (3.49), though the t are now identical across firms. From before, we also have the following optimal pricing equation and derivative:

$$p_n(0) = -\frac{t}{\phi[2+(N-1)t^2]}(\hat{\boldsymbol{b}}'\boldsymbol{d}) \quad (3.63)$$

$$\left.\frac{\partial p_n^*(\theta)}{\partial\theta}\right|_{\theta=0} = -\frac{t(2+(3N-2)t^2+(N-1)^2t^4)}{\phi(2+(N-1)t^2)(2+(2N-1)t^2)}(\hat{\boldsymbol{b}}'\boldsymbol{h}) \quad (3.64)$$

and optimal quantity equation and derivative:

$$q_n(\boldsymbol{p}^*, \theta) = -\phi\Omega_{nn}(\theta)p_n(\theta) \qquad \frac{\partial q_n(\boldsymbol{p}^*, \theta)}{\partial \theta}\bigg|_{\theta=0} = -\phi\bigg(\frac{1+(N-1)t^2}{1+Nt^2}\bigg)\frac{\partial p_n^*(\theta)}{\partial \theta}\bigg|_{\theta=0} \tag{3.65}$$

Lastly, note:

$$\frac{\partial}{\partial \theta}\frac{1}{2}\boldsymbol{g}_n(\theta)'\hat{C}\boldsymbol{g}_n(\theta)\bigg|_{\theta=0} = t^2\boldsymbol{h}'\hat{C}\boldsymbol{d} \tag{3.66}$$

These are useful for computing the impact of changing attribute orientation by one firm on its own expected profits:

$$\begin{aligned}
\frac{\partial \mathbb{E}\pi_n}{\partial \theta}\bigg|_{\theta=0} &= t\frac{\partial g(\theta)}{\partial \theta}\bigg|_{\theta=0}\bigg(p_n(0)\frac{q_n(\boldsymbol{p}^*, \theta)}{\partial \theta}\bigg|_{\theta=0} + q_n(\boldsymbol{p}^*, 0)\frac{\partial p_n^*(\theta)}{\partial \theta}\bigg|_{\theta=0} - t\hat{C}\boldsymbol{d}\bigg) \\
&= t\boldsymbol{h}'\bigg(-2\phi\bigg(\frac{1+(N-1)t^2}{1+Nt^2}\bigg)p_n(0)\frac{\partial p_n^*(\theta)}{\partial \theta}\bigg|_{\theta=0} - t\hat{C}\boldsymbol{d}\bigg) \\
&= -t^2\frac{2(1+(N-1)t^2)(2+(3N-2)t^2+(N-1)^2t^4)}{\phi(1+Nt^2)(2+(N-1)t^2)^2(2+(2N-1)t^2)}(\hat{\boldsymbol{b}}'\boldsymbol{d})(\hat{\boldsymbol{b}}'\boldsymbol{h}) - t^2\boldsymbol{h}'\hat{C}\boldsymbol{d} \\
&= t^2(K_n(\hat{\boldsymbol{b}}'\boldsymbol{d})(\hat{\boldsymbol{b}}'\boldsymbol{h}) - \boldsymbol{h}'\hat{C}\boldsymbol{d})
\end{aligned} \tag{3.67}$$

for $K_N = -\frac{2(1+(N-1)t^2)(2+(3N-2)t^2+(N-1)^2t^4)}{\phi(1+Nt^2)(2+(N-1)t^2)^2(2+(2N-1)t^2)}$. Thus, the symmetric equilibrium orientation $\boldsymbol{d}$ will be a stationary point if and only if:

$$\boldsymbol{h}'(K_N(\hat{\boldsymbol{b}}'\boldsymbol{d})\hat{\boldsymbol{b}} - \hat{C}\boldsymbol{d}) = 0 \qquad \boldsymbol{h} \perp \boldsymbol{d} \tag{3.68}$$

This is equivalent to the existence of some scalar μ such that:

$$\mu\boldsymbol{d} = K_N(\hat{\boldsymbol{b}}'\boldsymbol{d})\hat{\boldsymbol{b}} - \hat{C}\boldsymbol{d} \tag{3.69}$$

which we can re-arrange by multiplying both sides by $\boldsymbol{d}'$ and applying the optimisation condition for t^*:

$$\begin{aligned} \mu \boldsymbol{d} &= K_N(\hat{\boldsymbol{b}}'\boldsymbol{d})\hat{\boldsymbol{b}} - \hat{C}\boldsymbol{d} \\ \Leftrightarrow \mu &= K_N(\hat{\boldsymbol{b}}'\boldsymbol{d})^2 - \boldsymbol{d}'\hat{C}\boldsymbol{d} \\ \Leftrightarrow \mu &= (K_N - F_N)(\hat{\boldsymbol{b}}'\boldsymbol{d})^2 \end{aligned} \tag{3.70}$$

It can be shown that, for $N \geq 2$, $K_N > F_N > 0$ for all $t > 0$, meaning that, for the purpose of this setting, $\mu > 0$. We can re-arrange the initial equation to find the optimal orientation:

$$\begin{aligned} \mu \boldsymbol{d} &= K_N(\hat{\boldsymbol{b}}'\boldsymbol{d})\hat{\boldsymbol{b}} - \hat{C}\boldsymbol{d} \\ \Leftrightarrow (\hat{C} + \mu I)\boldsymbol{d} &= K_N(\hat{\boldsymbol{b}}'\boldsymbol{d})\hat{\boldsymbol{b}} \\ \Leftrightarrow \boldsymbol{d}^* &= K_N(\hat{\boldsymbol{b}}'\boldsymbol{d})(\hat{C} + \mu I)^{-1}\hat{\boldsymbol{b}} \\ \Rightarrow \boldsymbol{d}^* &\mid\mid (\hat{C} + \mu I)^{-1}\hat{\boldsymbol{b}} \\ \Rightarrow \boldsymbol{d}^* &\mid\mid \Gamma^{1/2}(C + \mu\Gamma)^{-1}\boldsymbol{b} \end{aligned} \tag{3.71}$$

If $N = 1$, as in the monopoly case, $\mu = 0$, yielding that original result. While we have not proved a global optimum, we have found a unique symmetric stationary equilibrium candidate; there is at most one interior symmetric equilibrium.

It may appear that this symmetric equilibrium only implies that firm's orientations are parallel, not that they are positively proportional. However, two goods cannot coexist with negatively proportional $\boldsymbol{d}$; if so, for at least one $\delta_n < 0$. The orientation rule above therefore pins down positive proportionality. This concludes the proof. □

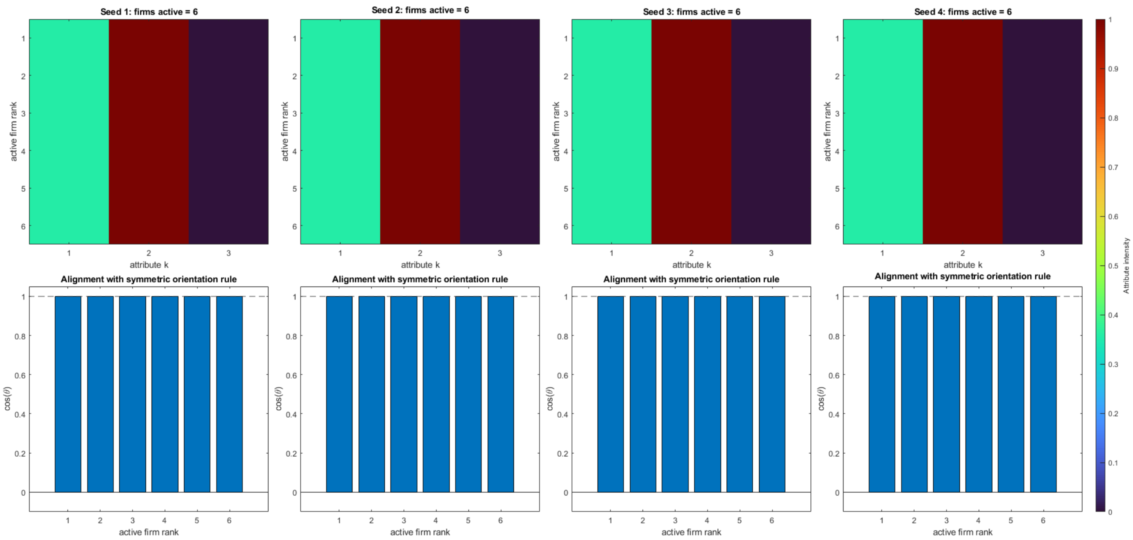
3.5.3 Discussion

The intuition behind the attribute exclusivity result is relatively straightforward. If firms optimise only over exclusive attributes, we have a diagonal M - goods are independent and there are no cross-price effects. Since each firm's design problem only affects product utility and demand slope, firms distribute their products attributes symmetrically following the same orientation rule as we saw in the monopoly case - though applied to their own exclusive attributes. This is a farily unique finding: two firms, in the same product category, with similar characteristics, ultimately act like monopolists - their strategic design levers are independent, and they make independent decisions. When a firm changes its exclusive attribute, it does not create a design-mediated cross-price effect on rivals. The mechanism can best be understood in terms of commitment, e.g., firms committing to product design optimisation in a way that minimises inter-comparability. This way, firms stop being close substitutes on the margins they can actually move.

Attribute non-exclusivity has its own unique implications. Because the inverted term will be relatively larger than the original $\hat{C}^{-1}$ for the lower-cost directions, this appears to favour greater dispersion of attributes per good than e.g. the multi-product monopoly and the attribute-exclusive single-product firms settings. Compared with a monopolist (in product or attributes), competitive firms tilt their attribute vector in all directions and spread themselves across greater attribute dimensions, proportionally to their salience. Competition makes firms more balanced in the attributes they load. The mechanism behind this is as follows: μ acts as a uniform penalty on concentrating too hard in the otherwise most favorable directions. In effect, it mirrors a greater competitive pressure, which pushes firms away in an attempt to minimise it along the most cost-effective popular attributes. Note how $K_N - F_N$ increases and then plateaus as the number of firms N increases - suggesting that greater competition biases consumers increasingly away from the attributes for which their costs are lower and towards those with lesser competitive salience.

For the attribute orientation case, we have only identified $\boldsymbol{s}^* = \boldsymbol{t}^* \boldsymbol{d}^*$ as a stationary point. Going beyond this is impractical and generally intractable. Via numerical simulation, we can however obtain a greater degree of confidence that this is indeed the unique symmetric equilibrium - and in fact is the only fixed point as a result of a convergence in iterated best-responses from different stating points. The graph below reflects such a numerical simulation, whereby firms converge to the symmetric equilibrium we have identified.

Figure 3.1: Selected sample attribute choice simulations under symmetric costs



Notes: Heatmap colours defines attribute intensity; alignment measured via cosine between attribute orientation and proposed orientation rule. Simulations converged after 113, 123, 131, and 130 iterations.

This doubles down on our finding that firms, even where faced with competitive pressure, still behave symmetrically, such that horizontal differentiation does not arise. Yet this may be because we have introduced no source of firm-/product-level heterogeneity. We consider this in the next section.

3.6 Firm- and Product-level Heterogeneity in Attribute Costs

So far, we have analysed a general case where firms decide upon different attributes with different costs. Nonetheless, we have so far excluded varying attribute costs per good/firm. The more complex version of this is a setting where:

$$\frac{1}{2}\boldsymbol{x}_n'\Sigma_n\boldsymbol{x}_n = \frac{1}{2}\boldsymbol{s}_n'(U^{-1}R\Sigma_n R'U)\boldsymbol{s}_n = \frac{1}{2}\boldsymbol{s}_n'C_n\boldsymbol{s}_n \tag{3.72}$$

This induces some interesting outcomes to both single-product firms and multi-product monopolists. Let us consider each in turn. For single-product firms, I will show the following:

Proposition 3.7: *Under firm-level heterogeneity in attribute costs, single-product firms optimising the design of their products given multiple attributes pursue horizontal differentiation, adjusting their designs both to weigh cost-weighted attribute utility and the strategic design of their competitors.*

Proof: One simple way to see how this affects single-product firms is to consider the FOC for a given firm n in the single-product firms setting under attribute non-exclusivity:

$$C_n\boldsymbol{s}_n = \frac{\partial \mathbb{E}\pi_n^*}{\partial \delta_n}\boldsymbol{b} + 2\frac{\partial \mathbb{E}\pi_n^*}{\partial M_{nn}}\Gamma\boldsymbol{s}_n + 2\sum_{i\neq n}^{N}\frac{\partial \mathbb{E}\pi_n^*}{\partial M_{ni}}\Gamma\boldsymbol{s}_i \tag{3.73}$$

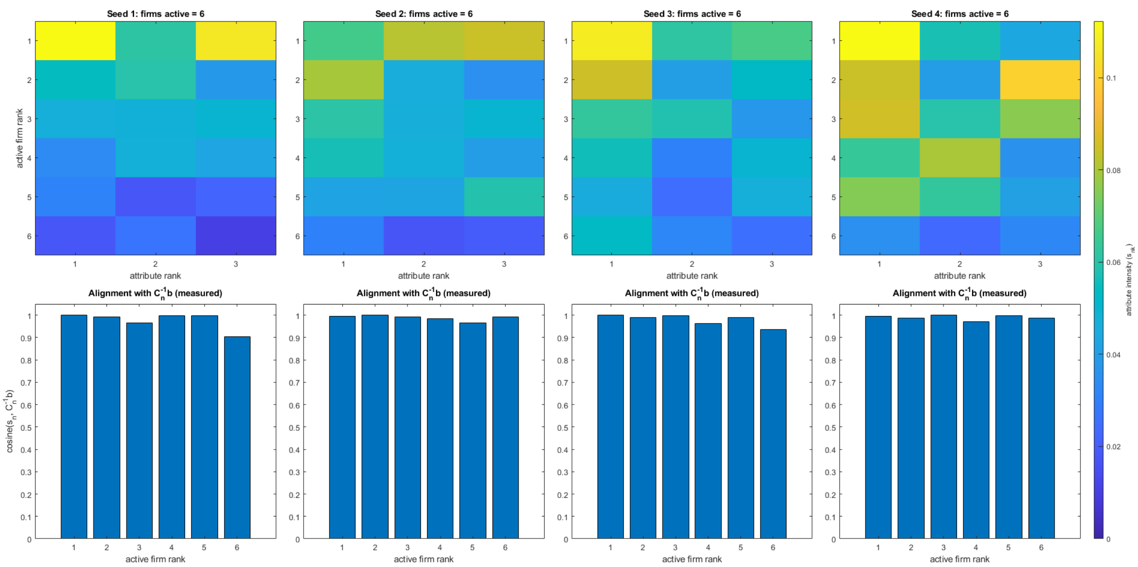
This completes the proof. □

Note how, as $M \to I$, the second and third terms, describing a strategic incentives towards/away from differentiation, go away. We can then write, for $\alpha_n > 0$ (since, by (3.2) and $M \to I$, $\partial\mathbb{E}\pi_n^*/\partial\delta_n > 0$):

$$C_n\boldsymbol{s}_n = \alpha_n\boldsymbol{b} \qquad \Leftrightarrow \qquad C_1\boldsymbol{s}_1 \;||\; \dots \;||\; C_N\boldsymbol{s}_N \quad \& \quad \boldsymbol{s}_n || C_n^{-1}\boldsymbol{b}, \quad \forall n = 1,\dots,N \tag{3.74}$$

Numerical assessments confirm a tendency towards this adjusted orientation rule:

Figure 3.2: Selected sample attribute choice simulations under asymmetric costs



Notes: Heatmap colours defines attribute intensity; alignment measured via cosine between attribute orientation and proposed orientation rule. Simulations converged after 826, 341, 1121, and 321 iterations.

The multi-product monopolist outcome faces unique changes as a result of heterogeneous attribute costs. I will prove the following:

Proposition 3.8*: Under good-level heterogeneity in attribute costs, a multi-product product-designing monopolist optimising its products given multiple attributes will sell only a single good; that which, given the attribute costs it faces, yields the greatest expected profit. No horizontal or vertical differentiation takes place.*

Proof: Let attribute costs differ by good, rather than firm as before. Let us set the per-good attribute costs in terms of $\boldsymbol{s}_k$ first. As before:

$$\frac{1}{2}\sum_{n=1}^{N}\boldsymbol{s}_n' C_n \boldsymbol{s}_n = \frac{1}{2}\sum_{k=1}^{K}\sum_{l=1}^{K}\boldsymbol{s}_k' D_{kl}\boldsymbol{s}_l \tag{3.75}$$

with $D_{kl}, \forall k, l = 1, \dots K$ a diagonal matrix which places the $C_{n,kl}$ element in the n-th diagonal. Then, from the expected profit function:

$$\frac{\partial \mathbb{E}\Pi}{\partial \boldsymbol{s}_k} = -\frac{1}{2\phi}(b_k - \gamma_k \boldsymbol{s}_k' M^{-1}\boldsymbol{\delta})M^{-1}\boldsymbol{\delta} - D_{kk}\boldsymbol{s}_k - \sum_{l\neq k}^{K} D_{kl}\boldsymbol{s}_l = 0 \tag{3.76}$$

For $\lambda_k = -\frac{1}{2\phi}(b_k - \gamma_k \boldsymbol{s}_k' M^{-1}\boldsymbol{\delta}),\ \forall k$, the following optimal orientation rule is then true:

$$\sum_{l=1}^{K} D_{kl}\boldsymbol{s}_l = \lambda_k \boldsymbol{y} \qquad \Leftrightarrow \qquad C_n \boldsymbol{s}_n = y_n \boldsymbol{\lambda}, \qquad \forall k = 1,\ldots,K, \quad n = 1\ldots,N \tag{3.77}$$

with $\boldsymbol{\lambda} = (\lambda_1\ \ldots\ \lambda_K)'$ and $\boldsymbol{y} = M^{-1}\boldsymbol{\delta}$. Row by row, the following equivalence arises:

$$C_n \boldsymbol{s}_n = y_n \boldsymbol{\lambda} \qquad \Rightarrow \qquad C_1\boldsymbol{s}_1\ ||\ \ldots\ ||\ C_N\boldsymbol{s}_N, \quad \forall n = 1,\ldots,N \tag{3.78}$$

This at first may appear to suggest that multi-product monopolists may deliver horizontal differentiation. Yet additional calculations reveal a different outcome. Note first that the LHS of the first equivalence can be re-written as follows:

$$\sum_{l=1}^{K} D_{kl}\boldsymbol{s}_l = \lambda_k \boldsymbol{y} \qquad \Leftrightarrow \qquad \boldsymbol{s}_k = A_k(\boldsymbol{\lambda})\boldsymbol{y} \tag{3.79}$$

for $A_k(\boldsymbol{\lambda})$ a diagonal matrix with diagonal elements $\{C_1^{-1}\boldsymbol{\lambda}\}_k,\ \ldots,\ \{C_N^{-1}\boldsymbol{\lambda}\}_k$, $\forall k =$

$1, \ldots, K$. Then, the following must be true:

$$\begin{aligned}
&\boldsymbol{\delta} = S\boldsymbol{b} \\
&\Leftrightarrow \boldsymbol{\delta} = \sum_{k=1}^{K} b_k \boldsymbol{s}_k \\
&\Leftrightarrow \boldsymbol{\delta} = \sum_{k=1}^{K} b_k A_k(\boldsymbol{\lambda})\boldsymbol{y} \\
&\Leftrightarrow M\boldsymbol{y} = \sum_{k=1}^{K} b_k A_k(\boldsymbol{\lambda})\boldsymbol{y} \\
&\Leftrightarrow (I + \sum_{k=1}^{K} \gamma_k \boldsymbol{s}_k \boldsymbol{s}_k')\boldsymbol{y} = \sum_{k=1}^{K} b_k A_k(\boldsymbol{\lambda})\boldsymbol{y} \\
&\Leftrightarrow \boldsymbol{y} + \sum_{k=1}^{K} \gamma_k \boldsymbol{s}_k (\boldsymbol{s}_k' \boldsymbol{y}) = \sum_{k=1}^{K} b_k A_k(\boldsymbol{\lambda})\boldsymbol{y} \\
&\Leftrightarrow \boldsymbol{y} + \sum_{k=1}^{K} \gamma_k (\boldsymbol{y}' A_k(\boldsymbol{\lambda})\boldsymbol{y}) A_k(\boldsymbol{\lambda})\boldsymbol{y} = \sum_{k=1}^{K} b_k A_k(\boldsymbol{\lambda})\boldsymbol{y}
\end{aligned} \tag{3.80}$$

Keep in mind that, from the earlier definition for λ_k, $\forall k = 1, \ldots, K$:$b_k - \gamma_k \boldsymbol{y}' A_k(\boldsymbol{\lambda})\boldsymbol{y} = b_k - \gamma_k \boldsymbol{s}_k' \boldsymbol{y} = -2\phi\lambda_k$. Then:

$$\begin{aligned}
&1 + \sum_{k=1}^{K} \gamma_k \{C_n^{-1}\boldsymbol{\lambda}\}_k (\boldsymbol{y}' A_k(\boldsymbol{\lambda})\boldsymbol{y}) = \sum_{k=1}^{K} b_k \{C_n^{-1}\boldsymbol{\lambda}\}_k \\
&\Leftrightarrow \sum_{k=1}^{K} (b_k - \gamma_k (\boldsymbol{y}' A_k(\boldsymbol{\lambda})\boldsymbol{y})) \{C_n^{-1}\boldsymbol{\lambda}\}_k = 1 \\
&\Leftrightarrow \sum_{k=1}^{K} \lambda_k \{C_n^{-1}\boldsymbol{\lambda}\}_k = \boldsymbol{\lambda}' C_n^{-1} \boldsymbol{\lambda} = -\frac{1}{2\phi}
\end{aligned} \tag{3.81}$$

Notably, this condition will only be satisfied for two goods simultaneously if:

$$\boldsymbol{\lambda}'(C_n^{-1} - C_m^{-1})\boldsymbol{\lambda} = 0 \tag{3.82}$$

This is ultimately a razor-thin margin. Hence, in general, only one good will satisfy it, meaning that the monopolist will only sell one good. Which good that will be will depend on which, if chosen, maximises expected profits. Heterogeneous costs have effectively introduced an incentive that breaks the indifference as to the distribution of attributes across goods sold by the monopolist. No horizontal differentiation takes place. □

Lastly, note how, under firm- and product-level heterogeneous costs, as $M \to I$, $\boldsymbol{\lambda} = \boldsymbol{b} - \Gamma S' M \boldsymbol{\delta} \approx \boldsymbol{b}$, matching the approximate orientation rule which arises from the single-product firms case under the same conditions.

3.7 Product Entry and Market Outcomes

To complete this Chapter, I tackle the matter of market outcome prediction for hypothetical new goods. This section does not provide an exhaustive exploration of product entry, but instead aims to respond directly to the point raised in Gandhi and Nevo (2021) regarding the ability of product-based models such as linear demand to predict demand from new goods. We will prove the following:

Proposition 3.10: *Greater product similarity between an incumbent firm and an entrant lowers incumbent firm demand and profits; makes a potential entrant less likely to enter; and has a minimal effect on consumer surplus. Upon entry, the incumbent's optimal product design tilts away from the monopoly orientation rule and loads more heavily on the exact opposite strategic direction to that determined by the entrant's design.*

Proof: Consider an incumbent and a new entrant, both single-product firms. Mathematically, we can show that demand for the incumbent's good is as follows:

$$q_{Inc} = \frac{m_{Ent}[(2m_{Inc}m_{Ent} - c^2)\delta_{Inc} - cm_1\delta_{Ent}]}{(m_{Inc}m_{Ent} - c^2)(4m_{Inc}m_{Ent} - c^2)} \tag{3.83}$$

for $m_i = 1 + \boldsymbol{s}_i'\Gamma\boldsymbol{s}_i$ and $c = \boldsymbol{s}_{Inc}'\Gamma\boldsymbol{s}_{Ent}$, where $M = [m_1\ c;\ c\ m_2]$. Expected firm profit on the other hand are as follows:

$$\mathbb{E}\pi_{Inc} = -\frac{m_{Ent}[(2m_{Inc}m_{Ent} - c^2)\delta_{Inc} - cm_1\delta_{Ent}]^2}{\phi(m_{Inc}m_{Ent} - c^2)(4m_{Inc}m_{Ent} - c^2)^2} \tag{3.84}$$

At $c = 0$, both firms behave as monopolists, and entry makes no difference. Consider instead the case where c is perturbed such that it reflects greater similarity in product characteristics. Then, incumbent

$$\frac{\partial q_{Inc}}{\partial c} \approx -\frac{\delta_{Ent}}{4m_{Inc}m_{Ent}} < 0 \tag{3.85}$$

$$\frac{\partial \mathbb{E}\pi_{Inc}}{\partial c} \approx \frac{\delta_{Inc}\delta_{Ent}}{\phi 4m_{Inc}m_{Ent}} < 0 \tag{3.86}$$

This suggests that around $c = 0$, greater characteristic overlap lowers both the incumbent firm' quantities and profits linearly. Zero-overlap entry is innocuous to the incumbent, while entry by a similar firm has a negative first-order effect that is linear in similarity. It can be shown that similarity has the identical effect for the entrant, meaning that similarity with incumbent goods makes entry less likely.

What about consumer welfare? It can be shown that similarity plays a minimal role, with effects driven primarily by initial product utility (much like Logit, consumer surplus increases by construction with product entry).

$$CS = \frac{1}{2}\boldsymbol{q}^{*'} M^{-1}\boldsymbol{q}^* \qquad \Rightarrow \qquad \frac{\partial}{\partial c}CS(c)\bigg|_{c=0} \approx 0 \tag{3.87}$$

This completes our proof. □

The model we have discussed in this Chapter also allows for further discussion on how new products, upon entry, affect optimal product design. From (3.73), pertaining to optimal attribute orientation for single-product firms under attribute cost heterogeneity, we know that optimal product design for good n, $\boldsymbol{s}_n^*$ is that such that $\boldsymbol{s}_n^*$ is a weighted combination of N directions: $\boldsymbol{b}$, as well as every other $\boldsymbol{s}_i$ vector. Furthermore, from the above, we know that $\frac{\partial \mathbb{E}\pi_{Inc}}{\partial M_{Inc,Ent}} = \frac{\partial \mathbb{E}\pi_{Inc}}{\partial c} < 0$. We would therefore expect entry to tilt design away from the monopoly direction $C^{-1}\boldsymbol{b}$ and toward a strategic direction determined by rival design. The negative sign implies that, for greater similarity between the incumbent and the new entrant's goods, the incumbent loads more heavily on desiging *away* from the entrant and less heavily towards $C^{-1}\boldsymbol{b}$. In other words, a new good will likely make its rivals simultaneously more opposed to itself and less like weighted attribute utility.

Much has been said in the product design literature on the concept of *niche* designs (see Bar-Isaac, Caruana, and Cunat, 2023, for a recent assessment). In particular, recent work has suggested that "extreme" designs may be an early strategy upon entry for a firm, with a greater focus on a broader audience over time reflecting something closer to equilibrium behaviour (Gong, 2021). This Chapter therefore provides one explanation - a niche design may make a firm more likely to enter a market, even if the equilibrium design is eventually of broader interest. While outside the scope of this Chapter, another way niche products these may arise in the model is if the assumption of a single representative consumer is broken, i.e. when consumers disagree as to what goods are substitutes and complements and to what extent. Further research is needed to determine if and when such degree of consumer heterogeneity takes place, since it does not necessarily follow from the existing literature on consumer segmentation/clustering.

Also outside the scope of this Chapter is a broader discussion on innovation. Innovation is often split between a new product (e.g. a new characteristic) or a new process (impacting production costs). It may be asserted that the model in this Chapter can help highlight another source of product innovation - a new technology mix. A new product, e.g. a phone with the same size but lighter, or an apple that stays ripe for longer without losing taste, changes the perceived relationship between product characteristics. For example, a new good which allows the same volume at much lower mass than previously observed changes the density distribution across the product category, impacting the design equilibrium. Further research should help uncover unique implications of this.

3.8 Numerical Example

To complete this Chapter, we briefly consider a working numerical example of the model described above. Consider a setting with two single-product price-setting firms, one selling a premium phone, and another a "budget" phone. These phones have two observable characteristics: battery life and camera quality. We can write the respective product characteristics matrix X as follows:

$$X = \begin{pmatrix} 20 & 16 \\ 10 & 7 \end{pmatrix} = \left(\frac{1}{\sqrt{5}} \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix} \right) \left(\sqrt{5} \begin{pmatrix} 10 & 7.8 \\ 0 & -0.4 \end{pmatrix} \right) = ZR \tag{3.88}$$

The premium phone has almost twice the "points" on both characteristics. "Points" here may refer to any unit we may wish to use for these characteristics: e.g. battery capacity in mAh; or camera resolution in megapixels. Let the former be the most relevant to a representative consumer:

$$\boldsymbol{\beta} = \begin{pmatrix} 0.10 \\ 0.05 \end{pmatrix} \quad \Rightarrow \quad \mathbb{E}\boldsymbol{\delta} = X\boldsymbol{\beta} = \begin{pmatrix} 2.8 \\ 1.35 \end{pmatrix} \tag{3.89}$$

We will also define the Hessian matrix of the representative consumer's utility function, M:

$$M = \begin{pmatrix} 3.5 & 1.5 \\ 1.5 & 3.5 \end{pmatrix} = (ZU)\Gamma(ZU)' + I = S\Gamma S' + I \tag{3.90}$$

$$S = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \; \Gamma = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}, \; U = \frac{1}{\sqrt{10}} \begin{pmatrix} 3 & 1 \\ -1 & 3 \end{pmatrix} \tag{3.91}$$

Despite the very different levels of their characteristics, to the point of being vertically differentiated in our observed characteristics, these goods appear to be imperfect substitutes with similar demand slopes. In re-writing M in terms of Γ and S, the model concilliates this paradox by weighing heavily the competitive salience of the first attribute in S, where both goods seem to load similarly, while weighing the competitive salience of the second attribute, where they differ more substantially, more lightly.

From the above, we can produce the implied attribute utility $\boldsymbol{b}$:

$$\boldsymbol{b} = S^{-1}\mathbb{E}\boldsymbol{\delta} = \begin{pmatrix} 2.93 \\ 1.03 \end{pmatrix} \tag{3.92}$$

To account for the significant differences in the number of points across the two goods, we pre-calculate their implied attribute costs as an "attribute budget". In practice, this means our results will resemble the setting with attribute cost heterogeneity. Let $\phi = -1$. The results are as follows:

$$X^*_{mono,excl} = \begin{bmatrix} 27.58 & 21.15 \\ 11.47 & 10.32 \end{bmatrix} \qquad X^*_{mono,non-excl} = \begin{bmatrix} 33.04 & 26.73 \\ 0 & 0 \end{bmatrix} \tag{3.93}$$

$$X^*_{sing,excl} = \begin{bmatrix} 27.58 & 21.15 \\ 11.47 & 10.32 \end{bmatrix} \qquad X^*_{sing,non-excl} = \begin{bmatrix} 30.98 & 25.07 \\ 11.39 & 9.20 \end{bmatrix} \tag{3.94}$$

The differences in cost across goods lead the monopoly towards the one-good outcome we discussed in a previous section; firm-level heterogeneity as a multiplicative of a common C on the other hand only affected attribute intensity - attributes are parallel as predicted by the cost homogeneity setup. The difference lies in that the monopoly is very sensitive to any heterogeneity, whereas individual firms require greater diversity of cost profiles to deliver greater variety of orientations.

Average expected profit per good are as follows:

$$\mathbb{E}\pi^*_{mono,non-excl} > \mathbb{E}\pi^*_{mono,excl} = \mathbb{E}\pi^*_{sing,excl} > \mathbb{E}\pi^*_{sing,non-excl} \tag{3.95}$$

$$\Leftrightarrow \; 0.3035 > 0.2733 = 0.2733 > 0.2212$$

As discussed in a previous section, outcomes are identical in the two attribute-exclusivity settings. The unique difference is that, for a monopolist, this is a constraint; for single-product firms, it constitutes an opportunity for collusion.

3.9 Discussion and Conclusion

This Chapter has shown how a fundamentally product-based demand system can be extended into a fully characteristics-based framework without losing the tractability that makes linear demand analytically useful. By allowing product characteristics to shape both consumers' valuations and the competitive relationships across goods, the Chapter demonstrates - in a precursor to future work - that linear demand can speak directly to questions on endogenous product design and entry. More broadly, the Chapter establishes that the model remains workable in settings with any finite number of goods, firms, and characteristics, and under both symmetry and asymmetry in cost structures.

I have not analysed the matter of consumer heterogeneity in this Chapter. It could be presumed this is due to some limitation of the model - this is not so. Once more than one consumer is considered, we move outside the realm of the simple proofs and closed-form solutions. Yet it has been convincingly demonstrated here that numerical solutions can converge fast in a linear demand model. Presumably, in a more complete model as in Chapter 1, firms could do as in Chapter 2 and study corner solutions, adjusting prices and characteristics to segment consumers. What this may look like mathematically would require extending this Chapter beyond what was its original scope: to demonstrate that linear demand is as robust an approach to study optimal product design and demand for new products as any existing alternative method.

References

Abaluck, J., and Adams, A. (2021). What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses. The Quarterly Journal of Economics, 136(3), 1611–1663.

Akerlof, G. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. The Quarterly Journal of Economics, 84(3), 488–500.

Allen, R., and Rehbeck, J. (2022). Latent complementarity in bundles models. Journal of Econometrics, 228(2), 322-341.

Amir, R., Erickson, P., and Jin J. (2017). On the microeconomic foundations of linear demand for differentiated products. Journal of Economic Theory, 169, 641-665.

Anderson, S., de Palma, A., and Thisse, J. (1992). Discrete Choice Theory of Product Differentiation. Cambridge, MA: The MIT Press.

Anderson, S., and Renault, R. (1999). Pricing, product diversity, and search costs: A Bertrand-Chamberlin-Diamond model". The RAND Journal of Economics, 30(4), 719–735.

Armstrong, T. (2016). Large market asymptotics for differentiated product demand estimators with economic models of supply. Econometrica, 84(5), 1961-1980.

Atalay, E., Frost, E., Sorensen, A., Sullivan, C., and Zhu, W. (2023). Scalable demand and markups (NBER Working Paper). National Bureau of Economic Research.

Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006). Who's Who in Networks. Wanted: The Key Player. Econometrica, 74(5), 1403-1417.

Backus, M., Conlon, C., and Sinkinson, M. (2021). Common Ownership in America: 1980–2017. American Economic Journal: Microeconomics, 13(3), 273-308.

Bar-Isaac, H., Caruana, G., and Cuñat, V. (2012). Search, Design, and Market Structure. American Economic Review, 102(2), 1140–60.

Bar-Isaac, H., Caruana, G., and Cuñat, V. (2023). Targeted Product Design. American Economic Journal: Microeconomics, 15(2), 157–86.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. Econometrica, 63(4), 841-890.

Blume, A. (2003). Bertrand without fudge. Economics Letters, 78(2), 167-168.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2013). Salience and Consumer Choice. Journal of Political Economy, 121(5), 803-843.

Campbell, J., and Viceira, L. (2002). Strategic Asset Allocation: Portfolio Choice for Long-Term Investors. Oxford: Oxford University Press.

Caplin, A., Dean, M., and Leahy, J. (2019). Rational inattention, optimal consideration sets, and stochastic choice. The Review of Economic Studies, 86(3), 1061–1094.

Caplin, A., Dean, M., and Martin, D. (2011). Search and Satisficing. American Economic Review, 101(7), 2011.

Carroll, C. (2000). Requiem for the Representative Consumer? Aggregate Implications of Microeconomic Consumption Behavior. American Economic Review, 90(2), 110–115.

Chen Y., and Riordan, M. (2007). Price and variety in the spokes model. The Economic Journal, 117(522), 897-921.

Cherchye, L., Crawford, I., de Rock, B., and Vermeulen, F. (2016). Gorman revisited: nonparametric conditions for exact linear aggregation. SERIEs: Journal of the Spanish Economic Association, 7, 203-220.

Choné, P., and Linnemer, L. (2020). Linear demand systems for differentiated goods: Overview and user's guide. International Journal of Industrial Organization, 73, 1-18.

Christensen, L., Jorgenson, D., and Lau, L. (1975). Transcendental Logarithmic Utility Functions. American Economic Review, 65(3), 367-383.

Crawford, G., Griffith, R., and Iaria, A. (2021). A survey of preference estimation with unobserved choice set heterogeneity," Journal of Econometrics, 222(1, A), 4-43.

Darton, R. (1980). Rotation in factor analysis. Journal of the Royal Statistical Society Series D: The Statistician, 29(3), 167–194.

d'Aspremont, C., Gabszewicz, J., and Thisse, J. (1979). On Hotelling's "stability in competition". Econometrica, 47(5), 1145–1150.

Deaton, A., and Muellbauer, J. (1980). An almost ideal demand system. American Economic Review, 70, 312-326.

Dixit, A., and Stiglitz, J. (1977). Monopolistic Competition and Optimum Product Diversity. American Economic Review, 67(3), 297-308.

Dopper, H., MacKay, A., Miller, N., and Stiebale, J. (2024). Rising Markups and the Role of Consumer Preferences. Journal of Political Economy, 133(8).

Draganska, M., Klapper, D., and Villas-Boas, S. (2010). A Larger Slice or a Larger Pie? An Empirical Investigation of Bargaining Power in the Distribution Channel. Marketing Science, 9(1), 57-74.

Dubé, J. (2005). Product Differentiation and Mergers in the Carbonated Soft Drink Industry. Journal of Economics and Management Strategy, 14(4), 879-904.

Eaton, B., and Lipsey, R. (1989). Product differentiation. In: Schmalensee, R.E., Willig, R.D. (Eds.), Handbook of industrial organization. North-Holland, Amsterdam, 723-768.

Economides, N. (1993). Quality variations in the circular model of variety-differentiated products. Regional Science and Urban Economics, 23(2), 235-257.

Eliaz, K., and Spiegler, R. (2011). Consideration Sets and Competitive Marketing. The Review of Economic Studies, 78(1), 235-262.

Ershov, D., Laliberté, J., Marcoux, M., and Orr, S. (2024). Estimating Complementarity with Large Choice Sets: an Application to Mergers. Available at SSRN: https://ssrn.com/abstract=3802097

Fama, E., and French, K. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1), 3-56.

Fosgerau, M., Monardo, J., and de Palma, A. (2024). The Inverse Product Differentiation Logit Model. American Economic Journal: Microeconomics, 16(4), 329-370.

Fox, T., and Lazzati, N. (2017). A note on identification of discrete choice models for bundles and binary games. Quantitative Economics, 8(3), 1021-1036.

Gandhi, A. and Houde, J. (2019). Measuring substitution patterns in differentiated-products industries. NBER Working Paper, 26375.

Gandhi, A., and Nevo, A. (2021). Empirical models of demand and supply in differentiated products industries. In K. Ho, A. Hortaçsu, and A. Lizzeri (Eds.), Handbook of Industrial Organization, 4(1), 63-139. Elsevier.

Gentzkow, M. (2007). Valuing new goods in a model with complementarity: Online newspapers. American Economic Review, 97(4), 713-744.

Sovinsky, M. (2008). Limited Information and Advertising in the U.S. Personal Computer Industry. Econometrica, 76(5), 1017-1074.

Golub, G., and van Loan, C. (1983). *Matrix computations.* Baltimore: Johns Hopkins University Press.

Gong, Z. (2021). Growing Influence. Working Paper.

Gorman, W. (1953). Community preference fields. Econometrica, 21(1), 63–80.

Gorman, W. M. (1961). On a class of preference fields. Metroeconomica, 13(2), 53–56.

Hauser, J. (2014). Consideration-set Heuristics. Journal of Business Research, 67(8), 1688-1699.

Hauser, J. and Wernerfelt, B. (1990). An Evaluation Cost Model of Consideration Sets. The Journal of Consumer Research, 16(4), 393-408.

Hausman, J., Leonard, G., and Zona J. (1994). Competitive analysis with differentiated products. Annales d'Économie et de Statistique, 34, 159-180.

Hausman, J. (1996). Valuation of new goods under perfect and imperfect competition. In: The Economics of New Goods, 207-248. National Bureau of Economic Research, Inc.

Hildenbrand, W. (1983). On the "Law of Demand". Econometrica, 51(4), 997-1019.

Hoberg, G., and Phillips, G. (2016). Text-Based Network Industries and Endogenous Product Differentiation. Journal of Political Economy, 124(5), 1423-1465.

Hotelling, H. (1929). Stability in competition. The Economic Journal, 39(153), 41-57.

Hristakeva, S. (2022). Vertical Contracts with Endogenous Product Selection: An Empirical Analysis of Vendor Allowance Contracts. Journal of Political Economy, 130(12), 3202-3252.

Iaria, A., and Wang, A. (2019). Identification and Estimation of Demand for Bundles. Available at SSRN: https://ssrn.com/abstract=3458543

Johnson, J., and Myatt, D. (2006). On the Simple Economics of Advertising, Marketing, and Product Design. American Economic Review, 96(3), 756–784.

Kirin Beer, (2020). Global Beer Consumption by Country in 2019. Tokyo: Kirin Holdings Company, Limited.

Kirman, A. (1992). Whom or What Does the Representative Individual Represent?. Journal of Economic Perspectives, 6(2), 117–136.

Koch, I. (2013). Analysis of multivariate and high-dimensional data. Cambridge University Press.

Koijen, R. and Yogo, M. (2019). A Demand System Approach to Asset Pricing. Journal of Political Economy, 127(4), 1475-1515.

Kreps, D., and Scheinkman, J. (1983). Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes. Bell Journal of Economics, 14(2), 326-337.

Lancaster, K. (1966). A new approach to consumer theory. Journal of Political Economy, 74, 132-157.

Lancaster, K. (1990). The economics of product variety: A survey. Marketing Science, 9(3), 189–206.

Lanier, J., Large, J., and Quah, J. (2023). Estimating very large demand systems. Available at https://ora.ox.ac.uk/objects/uuid:13accd61-5e96-46e8-849a-fdd7c627cbfd

Larson, N. (2013). Niche products, generic products, and consumer search. Economic Theory, 52(2), 793-832.

Lee, Y. (2024). Shrinkflation: Evidence on Product Downsizing and Consumer Response. Kilts Center at Chicago Booth Marketing Data Center Paper.

Lewbel, A., and Nesheim, L. (2019). Sparse demand systems: corners and complements. Boston College Working Papers in Economics 1005, Boston College Department of Economics.

Lin, B., Variyam, J., Allshouse, J., and Cromartie, J. (2003). Food and agricultural commodity consumption in the United States: Looking ahead to 2020 (Agricultural Economic Report No. 820). U.S. Department of Agriculture, Economic Research Service.

Lopes, C., Torres, D., Oliveira, A., Severo, M., Alarcão, V., Guiomar, S., Mota, J., Teixeira, P., Ramos, E., Rodrigues, S., Vilela, S., Oliveira, L., Nicola, P., Soares, S., Andersen. L (2017). Inquérito Alimentar Nacional e de Atividade Física (IAN-AF) 2015-2016. Universidade do Porto.

Magnolfi, L., McClure, J., and Sorensen, A. (2025). Triplet embeddings for demand estimation. American Economic Journal: Microeconomics, 17(1), 282-307.

Manchanda, P., Ansari, A., and Gupta, S. (1999). The "Shopping Basket": A Model for Multicategory Purchase Incidence Decisions. Marketing Science, 18(2), 95-114.

Manzini, P., Mariotti, M., and Ülkü, L. (2019). Stochastic Complementarity. The Economic Journal, 129(619), 1343-1363.

Marramaque, M., and Cardoso, F. (2021). Dairy Focus 1/2021: Portuguese EU Presidency. European Dairy Association.

Mas-Colell, A., Whinston, M., and Green, J. (1995). Microeconomic Theory. Oxford: Oxford University Press.

Matějka, F., and MacKay, A. (2015). Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model. American Economic Review, 105(1), 272-98.

McFadden, D. (2001). Economic Choices. American Economic Review, 91(3), 351-378.

McFadden, D. (1973) Conditional Logit Analysis of Qualitative Choice Behavior. In: Zarembka, P., Ed., Frontiers in Econometrics, Academic Press, 105-142.

Melitz, M., and Ottaviano, G. (2008). Market Size, Trade, and Productivity. The Review of

Economic Studies, 75(1), 295–316.

Miller, N., and Weinberg, M. (2017). Understanding the Price Effects of the Miller-Coors Joint Venture. Econometrica, 85(6), 1763-1791.

Miyashita, M. (2026). Characteristics Design:A Hedonic Approach to Optimal Product Differentiation. Working paper.

Moore, E. (1920). On the reciprocal of the general algebraic matrix. Bulletin of the American Mathematical Society, 26(9), 394-395.

Moorthy, K. (1988). Product and Price Competition in a Duopoly. Marketing Science, 7(2), 141-168.

Motta, M. (1993). Endogenous Quality Choice: Price vs. Quantity Competition. The Journal of Industrial Economics, 41(2), 113-131.

Mussa, M., and Rosen, S. (1978). Monopoly and product quality. Journal of Economic Theory, 18(2), 301-317.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. Econometrica, 69(2), 307-342.

Nocke, V., and Schutz, N. (2025). Oligopoly, Complementarities, and Transformed Potentials. CEPR Discussion Paper No. 20527. CEPR Press, Paris and London.

Peitz, M. and Sato, S. (2025). Asymmetric Platform Oligopoly. The RAND Journal of Economics, 57(1), 60-77.

Pellegrino, B. (2025). Product differentiation and oligopoly: A network approach. American Economic Review, 115(4), 1170-1225.

Penrose, R. (1955). A generalized inverse for matrices. Proceedings of the Cambridge Philosophical Society. 51(3), 406-413.

Perloff, J., and Salop, S. (1985). Equilibrium with Product Differentiation. The Review of Economic Studies, 52(1), 107-120.

Pinkse, J., and Slade, M. (2004). Mergers, brand competition, and the price of a pint. European Economic Review, 48(3), 617-643.

Rhodes, A., and Zhou, J. (2024). Personalized Pricing and Competition. American Economic Review, 114(7), 2141–70.

Ross, S. (1977). The Determination of Financial Structure: The Incentive-Signalling Approach. The Bell Journal of Economics, 8(1), 23-40.

Salop, S. (1979). Monopolistic competition with outside goods. The Bell Journal of

Economics, 10(1), 141-156.

Sebastian, R., Hoy, M., Murayi, T., Goldman, J., and Moshfegh, A. (2024). Breakfast Consumption by U.S. Adults: What We Eat in America, NHANES 2017 – March 2020 (Dietary Data Brief No. 59). U.S. Department of Agriculture, Agricultural Research Service.

Shaked, A., and Sutton, J. (1982). Relaxing price competition through product differentiation. The Review of Economic Studies, 49(1), 3–13.

Sherman, J., and Morrison, W. (1949). Adjustment of an Inverse Matrix Corresponding to Changes in the Elements of a Given Column or a Given Row of the Original Matrix. Annals of Mathematical Statistics. 20, 621.

Simon, H. (1957). Models of Man: Social and Rational. New York: John Wiley and Sons.

Spence, M. (1976). Product Selection, Fixed Costs, and Monopolistic Competition. The Review of Economic Studies, 43(2), 217-235.

Theil, H. (1965). The Information Approach to Demand Analysis. Econometrica, 33(1), 67-87.

Tian, Y., Lautz, S., Wallis, A., and Lambiotte, R. (2021). Extracting complements and substitutes from sales data: A network perspective. EPJ Data Science, 10(45).

Ushchev, P., and Zenou, Y. (2018). Price competition in product variety networks. Games and Economic Behavior, 110, 226-247.

Varian, H. (1980). A model of sales. The American Economic Review, 70(4), 651-659.

Vives, X. (1987). Small Income Effects: A Marshallian Theory of Consumer Surplus and Downward Sloping Demand. The Review of Economic Studies, 54(1), 87-103.

Voelkening, D. (2026). Product Design in an Hedonically Differentiated Duopoly. Working paper.

Von Ungern-Sternberg, T. (1988). Monopolistic Competition and General Purpose Products. The Review of Economic Studies, 55(2), 231-246.

Wales, T., and Woodland, A. (1983). Estimation of consumer demand systems with binding non-negativity constraints. Journal of Econometrics, 21(3), 263-285.

Wright P., and Barbour, F. (1977). Phased Decision Strategies: Sequels to an Initial Screening. Stanford School of Business Working Paper No. 353.

Wolinsky, A. (1986). True Monopolistic Competition as a Result of Imperfect Information. The Quarterly Journal of Economics, 101(3), 493-512.

Woodbury, M. (1950). Inverting modified matrices (Memorandum Rept. 42, Statistical

Research Group). Princeton: Princeton University.

Zeitlin, J., and Rangoni, B. (2025). How the European Union reconciles uniform regulation with legitimate diversity: towards a tighter experimentalist governance architecture. Journal of European Public Policy, 1–29.

Appendices

A

Appendix - Chapter 1

Contents

A.1 Proofs **131**

A.2 Welfare analysis **134**

A.3 Details on usage of transaction sample data **136**

A.4 Accounting for multi-product monopolistic competition . . **136**

A.5 Retail stock-outs **137**

A.1 Proofs

Lemma 1.1. By construction, $\boldsymbol{x}'M\boldsymbol{x} > 0$ for all non-zero $\boldsymbol{x}$. $\boldsymbol{y}'A'MA\boldsymbol{y} \geq 0$ because, where $A\boldsymbol{y} \neq 0$, $\boldsymbol{y}'A'MA\boldsymbol{y} = \boldsymbol{x}'M\boldsymbol{x} > 0$ for non-zero $\boldsymbol{x}$. If $A\boldsymbol{y} = 0$, $\boldsymbol{x}'M\boldsymbol{x} = 0$ because $\boldsymbol{x} = A\boldsymbol{y} = 0$. □

Lemma 1.2. From an individual consumer's optimisation in T_i, the stationarity condition is:

$$\hat{\boldsymbol{z}}_i' A_i' (\boldsymbol{\delta}_i + \phi_i \boldsymbol{p}) - T_i \hat{\boldsymbol{z}}_i' A_i' M_i A_i \hat{\boldsymbol{z}}_i = 0 \tag{A.1}$$

From (1.8), we have the stationarity condition from an individual consumer in $\hat{\boldsymbol{z}}_i$:

$$A_i'(\boldsymbol{\delta} + \phi_i \boldsymbol{p}) - T_i (A_i' M_i A_i) \hat{\boldsymbol{z}}_i + \frac{1}{T_i}(\hat{\boldsymbol{\lambda}}_{\boldsymbol{i}} + \eta_i \mathbf{1}) = 0 \tag{A.2}$$

If we multiply the latter by $\hat{\boldsymbol{z}}_i'$, we can take advantage of (i) the complementary slackness condition such that $\hat{\boldsymbol{z}}_i' \hat{\boldsymbol{\lambda}}_i = 0$, and (ii) the NORM constraint such that $\mathbf{1}'\boldsymbol{z} = 1$, so:

$$\hat{\boldsymbol{z}}_i' A_i' (\boldsymbol{\delta} + \phi_i \boldsymbol{p}) - T_i \hat{\boldsymbol{z}}_i' (A_i' M_i A_i) \hat{\boldsymbol{z}}_i + \eta_i / T_i = 0 \tag{A.3}$$

For both stationarity conditions to hold, we thus must have $\eta_i = 0$.

Note that, should we have an additional condition restricting some elements of $\hat{\boldsymbol{z}}_i$ to 0, the individual consumer's Lagrangean would include a zero-purchase multiplier, e.g. $\boldsymbol{\gamma}_i$. By construction, such a vector's elements would be positive only where $\hat{\boldsymbol{z}}_i$'s elements are equal to 0 and vice-versa, so such a term would drop out once multiplied by $\hat{\boldsymbol{z}}_i'$. The finding that $\eta_i = 0$ would therefore apply also under that more complex setting. □

Lemma 1.3. Regarding symmetry: for M positive definite and symmetric (by construction), $A'MA$ is symmetric - $(A'MA)' = A'M'A = A'MA$. By eigenvalue decomposition, $A'MA = U\Lambda U'$, and $(A'MA)^+ = U\Lambda^+ U'$, where Λ^+ is formed by replacing each non-zero $\lambda, \forall i$ with $1/\lambda_i$, and leaving zeros intact. Since Λ^+ is diagonal, it is symmetric, and therefore $(A'MA)^+$ is also symmetric. Lastly, if $(A'MA)^+$ is symmetric,

$A(A'MA)^+A'$ is also symmetric - $(A'(A'MA)^+A')' = A'((A'MA)^+)'A' = A(A'MA)^+A'$. Regarding positive semi-definiteness: If $A'MA$ is positive semi-definite, $(A'MA)^+$ is also positive semi-definite. As shown above, $A'MA$ is symmetric and can be diagonalised via eigenvalue decomposition, and the pseudoinverse inverts the non-zero eigenvalues while leaving the zero eigenvalues at zero. The inversion does not change the signs of the eigenvalues, so $\forall\lambda \geq 0$ still holds. If $(A'MA)^+$ is positive semi-definite, $\boldsymbol{x}'(A'MA)^+\boldsymbol{x} \geq 0$. This is the case for $\boldsymbol{x} = A'\boldsymbol{y} \neq 0$. For $\boldsymbol{x} = A'\boldsymbol{y} = 0$, $\boldsymbol{x}'(A'MA)^+\boldsymbol{x} = 0$. Therefore, $\boldsymbol{x}'(A'MA)^+\boldsymbol{x}' = \boldsymbol{y}'A(A'MA)^+A'\boldsymbol{y} \geq 0$ for all non-zero (and zero) $\boldsymbol{y}$. $\square$

Lemma 1.4. If A is full row rank, it can be shown that $A(A'MA)^+A' = M^{-1}$. Its singular value decomposition could be written as $A = U\Sigma V'$, for U a $K \times K$ orthogonal matrix; $\Sigma = [\Sigma_r 0]$ a $K \times J$ matrix, with Σ_r a $K \times K$ diagonal matrix containing the nonzero singular values of A, and 0 a $K \times (J-K)$ zero block; and $V = [V_1 V_2]$ is a $J \times J$ orthogonal matrix where V_1 is a $J \times K$ matrix whose columns form the orthonormal basis for the row space of A, while V_2 is a $J \times (J-K)$ matrix spanning the null space of A. A can then be rewritten as $A = U\Sigma_r V_1'$. From here, $A(A'MA)^+A' = U\Sigma_r V_1'(V_1\Sigma_r U'MU\Sigma_r V_1')^+V_1\Sigma_r U' = U\Sigma_r V_1'V_1(\Sigma_r U'MU\Sigma_r)^{-1}V_1'V_1\Sigma_r U' = U\Sigma_r(\Sigma_r U'MU\Sigma_r)^{-1}\Sigma_r U' = U(U'MU)^{-1}U' = M^{-1}$ by the properties of orthogonal and diagonal matrices. $\square$

Lemma 1.5. The sum of a positive definite with a positive semi-definite matrix is not always invertible. By theorem, however, the sum of a positive definite Ω with a positive semi-definite matrix $A'(A'MA)^+A'$, *where the positive definite matrix is diagonal with strictly positive diagonal elements*, is itself positive definite, and therefore invertible: $\boldsymbol{x}'\Omega + A(A'MA)^+A'\boldsymbol{x} = \boldsymbol{x}'\Omega\boldsymbol{x} + \boldsymbol{x}'A(A'MA)^+A'\boldsymbol{x} =$ positive + non-negative > 0. We can show that Ω is strictly positive in its diagonal. Express the diagonal matrix whose diagonal matches that of $A(A'MA)^+A'$ as Ω. To guarantee strictly positive elements in the diagonal of Ω, two requirements must be satisfied. Firstly, $\boldsymbol{v}_i \neq 0, \forall i$, for $\boldsymbol{v}_i = A'\boldsymbol{e}_i$, where $\boldsymbol{e}_i$ is the standard basis vector, and $\Omega_{ii} = \boldsymbol{e}_i'A(A'MA)^+A'\boldsymbol{e}_i = \boldsymbol{v}_i'(A'MA)^+\boldsymbol{v}_i$, where

$\boldsymbol{e}_i$ "picks out" the i-th diagonal of Ω. This is the case as long as row i of A has at least one non-zero term, which is true of every row in A by definition, as otherwise said row would describe a product we have not observed in the data. Secondly, $\boldsymbol{v}_i'(A'MA)^{+}\boldsymbol{v}_i > 0, \forall i$. Now, as we mentioned earlier, $(A'MA)^{+}$ is positive semi-definite. However, $\boldsymbol{v}_i$ is restricted to the column space of A by construction, the same column space of $(A'MA)^{+}$. On the subspace where $A'MA$ is "active" (its column space), the pseudoinverse $(A'MA)^{+}$ behaves like a true inverse and is positive definite on that subspace. This ensures that whenever is $\boldsymbol{v}_i$ is nonzero (i.e., when it lies in the column space of A), we have $\boldsymbol{v}_i'(A'MA)^{+}\boldsymbol{v}_i > 0, \forall i$. Given our proof of strictly positive diagonal elements in Ω, $\Omega + A(A'MA)^{+}A'$ is invertible. $\square$

A.2 Welfare analysis

Representation rationalisation is weaker than a full aggregation result, with implications for calculating social welfare. To see this, note:

$$\begin{aligned}
\sum_{i=1}^{N} U_i(\boldsymbol{Q}/N, W) &= \sum_{i=1}^{N}[(\boldsymbol{Q}/N)'(\boldsymbol{\delta}_i + \phi_i \boldsymbol{p}) - \frac{1}{2}(\boldsymbol{Q}/N)'M(\boldsymbol{Q}/N)] + \sum_{i=1}^{N}\phi_i W \\
&= (\boldsymbol{Q}/N)'(\boldsymbol{\delta} + \phi \boldsymbol{p}) - \sum_{i=1}^{N}\frac{1}{2}(\boldsymbol{Q}/N)'M(\boldsymbol{Q}/N) + \phi W \\
&= (\boldsymbol{Q}/N)'(\boldsymbol{\delta} + \phi \boldsymbol{p}) - N \cdot \frac{1}{2}(\boldsymbol{Q}/N)'M(\boldsymbol{Q}/N) + \phi W \\
&= (\boldsymbol{Q}/N)'(\boldsymbol{\delta} + \phi \boldsymbol{p}) - \frac{1}{2N}\boldsymbol{Q}'M\boldsymbol{Q} + \phi W \\
&= \frac{1}{N}U_{RC}(\boldsymbol{Q}, W)
\end{aligned} \tag{A.4}$$

for $W = \sum_{i=1}^{N} w_i$. Therefore:

$$U_{RC}(\boldsymbol{Q}, W) \propto \sum_{i=1}^{N} U_i(\boldsymbol{Q}/N, W) \neq \sum_{i=1}^{N} U_i(\boldsymbol{q}_i) \tag{A.5}$$

with proportionality also a result of utility functions being unique only up to a scale. Defining U_{RC} over the aggregates $\boldsymbol{\delta}$ and ϕ does not deliver aggregation at the utility level.

Up to an affine transformation, it can nonetheless be shown that $U_{RC}(\boldsymbol{Q})$ is proportional to the utilitarian planner's maximum attainable social welfare SW as a function of $\boldsymbol{Q}$:

$$\begin{aligned} SW(\boldsymbol{Q}) &= \max_{\{\boldsymbol{q}_i\}_{i=1}^N : \sum_{i=1}^N \boldsymbol{q}_i = \boldsymbol{Q}} \sum_{i=1}^N (\boldsymbol{q}_i' \boldsymbol{\delta}_i - \frac{1}{2} \boldsymbol{q}_i' M \boldsymbol{q}_i) \\ &\Leftrightarrow \boldsymbol{q}_i^*(\boldsymbol{Q}) = (\boldsymbol{Q}/N) + M^{-1}(\boldsymbol{\delta}_i - \bar{\boldsymbol{\delta}}), \ \forall i \end{aligned} \tag{A.6}$$

The planner gives everyone an equal split, then adjusts each consumer's allocation toward the goods for which that consumer has above-average initial utility (average being $\bar{\boldsymbol{\delta}}$). Then:

$$\begin{aligned} SW^*(\boldsymbol{Q}) &= \sum_{i=1}^N (\boldsymbol{q}_i^*(\boldsymbol{Q})' \boldsymbol{\delta}_i - \frac{1}{2} \boldsymbol{q}_i^*(\boldsymbol{Q})' M \boldsymbol{q}_i^*(\boldsymbol{Q})) \\ &= \boldsymbol{Q}' \bar{\boldsymbol{\delta}} - \frac{1}{2N} \boldsymbol{Q}' M \boldsymbol{Q} \ + \ \frac{1}{2} \sum_{i=1}^N (\boldsymbol{\delta}_i - \bar{\boldsymbol{\delta}})' M^{-1} (\boldsymbol{\delta}_i - \bar{\boldsymbol{\delta}}) \\ &= \sum_{i=1}^N U_i(\boldsymbol{Q}/N) + \frac{1}{2} \sum_{i=1}^N (\boldsymbol{\delta}_i - \bar{\boldsymbol{\delta}})' M^{-1} (\delta_i - \bar{\boldsymbol{\delta}}) \\ &= N \cdot U_{RC}(\boldsymbol{Q}) + [\text{constant in } \boldsymbol{Q} \geq 0] \end{aligned} \tag{A.7}$$

In other words, the utility obtained by a representative consumer with aggregate parameters $\boldsymbol{\delta}$ and ϕ can be used to study how changes in aggregate demand impact the social welfare obtained by a social planner finding the best achievable welfare via transfers. This makes the framework well-suited for social welfare comparisons across policies that operate primarily through changes in aggregate consumption $\boldsymbol{Q}$. This approach depends on interior solutions. If individual non-negativity constraints bind, as they do throughout the Chapter, the expression would need KKT multipliers.

A.3 Details on usage of transaction sample data

Because, by **Proposition 2** and the Minkowski sum, $\boldsymbol{z}$ is unique in $\text{col}(A)$, T is also unique, for $T = \sum_{i=1}^{N} T_i$. However, because only $\boldsymbol{Q}$ is identified and not $\boldsymbol{Q}/T$, T is not identified. Therefore, in practice, the only implication of T for an analysis of aggregated demand is that, should we want to estimate parameters for the average demand per transaction yet only observe a random sample of transactions T_r, we can divide the resulting $\boldsymbol{Q}_r$ by T_r to obtain:

$$\frac{1}{T_r}\boldsymbol{Q}_r = \frac{1}{T_r}(\frac{T_r}{T}\boldsymbol{Q}) = \frac{1}{T}\boldsymbol{Q} \tag{A.8}$$

This is our setting below, such that all estimated parameters must be understood as relative to the average transaction.

A.4 Accounting for multi-product monopolistic competition

In Chapter 1, we considered two extreme market structures: that (i) of a finite number of single-product firms, and (ii) of a single multi-product monopoly. In this Appendix, I extend the analysis to incorporate multi-product monopolistic competition, whereby simultaneous multi-product and single-product firms compete with each other. More specifically, I provide a condition under which a unique equilibrium can be computed for these markets. Discussion for the model's implications for entry dynamics is left for future work.

For a more general settings with multiple single- and multi-product firms, I propose imposing a bound on the extent to which cross-price effects from goods owned by the same multi-product firm are accounted for by firms:

$$W = \Omega + M^{-1} + \alpha(M^{-1} \circ G) \tag{A.9}$$

Then, if $M_{ii}^{-1} + \Omega_{ii} = 2M_{ii}^{-1} > \sum_{j \neq i: G_{ij}=0} |M_{ij}^{-1}| + (1+\alpha) \sum_{j \neq i: G_{ij}=1} |M_{ij}^{-1}|, \ \forall i$, or:

$$\alpha < \min_i \frac{2M_{ii}^{-1} - \sum_{j \neq i: G_{ij}=0} |M_{ij}^{-1}|}{\sum_{j \neq i: G_{ij}=1} |M_{ij}^{-1}|} - 1 \tag{A.10}$$

then the resulting demand Jacobian implies a well-behaved demand system.[1] Alternatively, we have a both sufficient and necessary condition in:

$$\alpha < \min_{\boldsymbol{x} \neq 0, \boldsymbol{x}'(M^{-1} \circ G)\boldsymbol{x} < 0} \frac{\boldsymbol{x}'(\Omega + M^{-1})\boldsymbol{x}}{-\boldsymbol{x}'(M^{-1} \circ G)\boldsymbol{x}} \tag{A.11}$$

This approach is less conservative but may be harder to calculate in some settings.

A.5 Retail stock-outs

Expressions derived from the consumer choice model in Chapter 1 can be particularly useful to understand the impact on retail settings where a given product goes out of stock. We return to the matter of price equilibria again at the end of the next section.

Lemma A.1*: Following the stock-out of a good sold by a monopolist retailer, the resulting changes in optimal demand, aggregate profits, and consumer surplus have the following closed-form expression:*

$$\Delta \boldsymbol{Q}^* = -\frac{1}{4} \begin{pmatrix} \sigma_{ii}\delta_i + \boldsymbol{\sigma}_i'\boldsymbol{\delta}_{-i} \\ \boldsymbol{\sigma}_i \delta_i \end{pmatrix} \tag{A.12}$$

[1] This is effectively a strict diagonal dominance condition.

$$\Delta\Pi = -\frac{1}{4\phi}\Big[\sigma_{ii}\delta_i^2 + 2\delta_i(\boldsymbol{\sigma}_i'\boldsymbol{\delta}_{-i})\Big] \tag{A.13}$$

$$\Delta CS = -\frac{3}{8}\Big[\sigma_{ii}\delta_i^2 + 2\delta_i(\boldsymbol{\sigma}_i'\boldsymbol{\delta}_{-i})\Big] \tag{A.14}$$

for $\Sigma = A(A'MA)^{+}A'$.

Proof: All model outputs are in one of the following forms: $\boldsymbol{y_1} = \alpha\boldsymbol{x}'\Sigma\boldsymbol{x}$ or $\boldsymbol{y_2} = \alpha\Sigma\boldsymbol{x}$ for α a strictly positive scalar, $\boldsymbol{x}$ a strictly positive vector, and Σ a positive semi-definite matrix. Suppose we wish to consider the impacts derived from the removal of the i-th component of $\boldsymbol{x}$ and Σ. These terms can be expressed as such:

$$\boldsymbol{x} = \begin{pmatrix} x_i \\ \boldsymbol{x}_{-i} \end{pmatrix} \tag{A.15}$$

and

$$\Sigma = \begin{pmatrix} \sigma_{ii} & \boldsymbol{\sigma}_i' \\ \boldsymbol{\sigma}_i & \Sigma_{-i} \end{pmatrix} \tag{A.16}$$

The notation for $\boldsymbol{\sigma}_i$ excludes the element corresponding to σ_{ii}. Starting with $\boldsymbol{y_1}$, we may restate it as:

$$\boldsymbol{y_1} = \sigma_{ii}x_i^2 + 2x_i(\boldsymbol{\sigma}_i'\boldsymbol{x}_{-i}) + \boldsymbol{x}_{-i}\Sigma{-i}\boldsymbol{x}_{-i} \tag{A.17}$$

of which the first two terms are the contribution from the i-th dimension of both $\boldsymbol{x}$ and Σ. If we remove the i-th product, i.e. the i-th dimension of $\boldsymbol{x}$ and Σ, the difference between the original and reduced term is:

$$\Delta_1 = -\alpha\Big[\sigma_{ii}x_i^2 + 2x_i(\boldsymbol{\sigma}_i'\boldsymbol{x}_{-i})\Big] \tag{A.18}$$

We can similarly re-write $\boldsymbol{y_2}$:

$$\boldsymbol{y_2} = \alpha \begin{pmatrix} \sigma_{ii} x_i + \boldsymbol{\sigma}_i' \boldsymbol{x}_{-i} \\ \boldsymbol{\sigma}_i x_i + \Sigma_{-i} \boldsymbol{x}_{-i} \end{pmatrix} \tag{A.19}$$

If the i-th component is dropped, the first (upper) block drops out, and the lower block's reduced version is $\alpha \Sigma_{-i} \boldsymbol{x}_{-i}$. □

Even where M is symmetric and diagonally dominant (common assumptions, sufficient for positive definiteness), consumer surplus and aggregate profits may or may not *decrease* as a result of product exit, and the ways in which they do so will depend on the structure of the representative consumer's consideration set. The QQUM specification aids us in allowing the expressions above to be written in closed form. This may not be the whole story however; a product whose removal raises aggregate profit may nonetheless be kept within the assortment if it can minimise the impact of other products' removal. As can be seen above, the removal of any given product is exacerbated or dampened by its interaction with others. A correct understanding of the interaction matrix is important for addressing stock-outs.

B

Appendix - Chapter 2

Contents

B.1 Handling less-than-full-row-rank consideration sets in empirical estimation . . . 141

B.2 Participation rates . . . 142

B.2.1 Transaction trends . . . 144

B.2.2 Estimation of participation rate elasticity . . . 145

B.1 Handling less-than-full-row-rank consideration sets in empirical estimation

What if the full NN-LF condition binds? We have assumed thus far that, for an appropriate proxy W, we have that $M^{-1} \approx \sum_{j=0}^{J} \alpha_j W^j$. What is less clear is whether W is an appropriate proxy for M^{-1} or if it should be used to approximate $A(A'MA)^{+}A'$ as a whole where the NN-LF binds. This is a matter better left to a test: model both cases and verify which fits the data best. This is less so a problem in the latter case, as our estimation approach above already fits that alternative without any additional changes. The former, however, requires expressing $A(A'MA)^{+}A'$ as a function of M^{-1}, so that $\sum_{j=0}^{J} \alpha_j W^j$ can be fit. We could use $M = I - \alpha W$ but the pseudo-inversion step would not be able to follow the Neumann series, introducing a bottleneck on estimation.

To achieve this, note:

Lemma B.1: *$A(A'MA)^{+}A' = M^{-1}A(A'M^{-1}A)^{+}A'M^{-1}$, for M a positive definite matrix.*

Proof: The intermediate step is to show that $A(A'MA)^{+}A'M = M^{-1}A(A'M^{-1}A)^{+}A'$; if this is the case, both sides of the equation can be multiplied by M^{-1} to yield the Lemma's equality. I do so by noting that both sides of the equation are an M-orthogonal projection onto the same space $col(A)$. Since orthogonal projections are unique, the two must be equivalent. An orthogonal projection to space $col(A)$ must satisfy three conditions: the matrix must be in $col(A)$, be idempotent, and self-adjoint in M.

For a vector $\boldsymbol{u} = A\boldsymbol{v} \in col(A)$, $A(A'MA)^{+}A'M(Av) = Av$. Similarly, noting that $col(M^{-1}A) = col(A)$ as M^{-1} is positive definite with $col(M^{-1}) = \mathbb{R}$, we have $\boldsymbol{u} = M^{-1}A\boldsymbol{v} \in col(A)$ such that $M^{-1}A(A'M^{-1}A)^{+}A'(M^{-1}Av) = M^{-1}Av$. This means that every vector in $col(A)$ is also in the column space of these expressions. The proof is complete by noting the reverse - the column space of these expressions is also a subset of $col(A)$: $A(A'MA)^{+}A'M = AB$ for $B = (A'MA)^{+}A'M$, and $col(AB) \subseteq col(A)$; $M^{-1}A(A'M^{-1}A)^{+}A' = M^{-1}AB$ for $B = (A'M^{-1}A)^{+}A'$, with $col(M^{-1}AB) \subseteq col(M^{-1}A) = col(A)$. Idempotency is similarly straightforward: $A(A'MA)^{+}A'M \cdot A(A'MA)^{+}A'M = A(A'MA)^{+}A'M$ and $M^{-1}A(A'M^{-1}A)^{+}A' \cdot M^{-1}A(A'M^{-1}A)^{+}A' = M^{-1}A(A'M^{-1}A)^{+}A'$. Lastly, note:

$$(A(A'MA)^{+}A'M\boldsymbol{x})'M\boldsymbol{y} = \boldsymbol{x}'M(A(A'MA)^{+}A'M\boldsymbol{y}) \tag{B.1}$$

and

$$(M^{-1}A(A'M^{-1}A)^{+}A'\boldsymbol{x})'M\boldsymbol{y} = \boldsymbol{x}'M(M^{-1}A(A'M^{-1}A)^{+}A'\boldsymbol{y}) \tag{B.2}$$

□

If $A(A'MA)^{+}A' = M^{-1}A(A'M^{-1}A)^{+}A'M^{-1}$, then we can implement grid search to jointly estimate the parameters in $M^{-1} \approx \sum_{j=0}^{J} \alpha_j W^j$: set initial values for $\alpha_j, \forall j$, calculate $\hat{M^{-1}}A(A'\hat{M^{-1}}A)^{+}A'\hat{M^{-1}}$, and regress the model to obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\phi}$. The optimal $\boldsymbol{\alpha}$ will be those which best fits the data.

B.2 Participation rates

In our empirical application, all stores were at least 1km away from the nearest store competitor at the beginning of the sample in 2020, except for Portimão (700m away). Maia faced entry by a competitor 700m away from its store in February 2023, just after the end of our sample, and another 1.4km away in late February 2022; Mafra faced (re-)entry 300m away in February 2021; and Portimão faced an additional entry 600m away in June 2022. This high level of activity in the industry reflects a recent expansion strategy pursued by

European discount retailers in the Portuguese market, which remained active throughout the COVID pandemic. Given this thesis's focus on intra-supermarket competition, these may raise concerns that inter-supermarket competition impacts our outcomes.

In addition, in the standardised regressions above, all product quantities purchased are corrected by the number of transactions observed each period in the relevant supermarket. This is in line with linear demand aggregation. However, such correction introduces endogeneity concerns: if prices at the grocery store increase, consumers may decide to shop elsewhere, and vice versa. What may be the implications for mark-ups? Start from the differentiated-product, supermarket-time specific, Bertrand-Nash, Lerner index:

$$L_{m,t} = -((G \circ E_{m,t})^{-1}) \cdot \mathbf{1} \tag{B.3}$$

where G is the ownership matrix, and

$$E_{m,t} = J_{m,t} \cdot p_{m,t}/s_{m,t} \tag{B.4}$$

with s the per-customer demand, and $J = \frac{\partial s}{\partial p}$, each specified per market m and time period t. If participation N varies with prices, the per-customer Jacobian we estimated earlier from demand, $\hat{J}_{m,t} = \frac{\partial s}{\partial p}$, bundles two margins:

$$\frac{\partial s}{\partial p} = (1/N) \cdot \frac{\partial q}{\partial p} - s \cdot (\eta^N)' \tag{B.5}$$

where $\eta_j^N \equiv \frac{\partial \ln N}{\partial p_j}$ (participation sensitivity). Thus the intensive-margin Jacobian needed for mark-ups is:

$$J^{int}_{m,t} = \hat{J}_{m,t} + s_{m,t} \cdot (\eta^N_{m,t})' \tag{B.6}$$

If $\frac{\partial q}{\partial p} < 0$ (higher prices reduce participation), ignoring participation would make own-price slopes look too small in magnitude, which would inflate the Lerner indices.

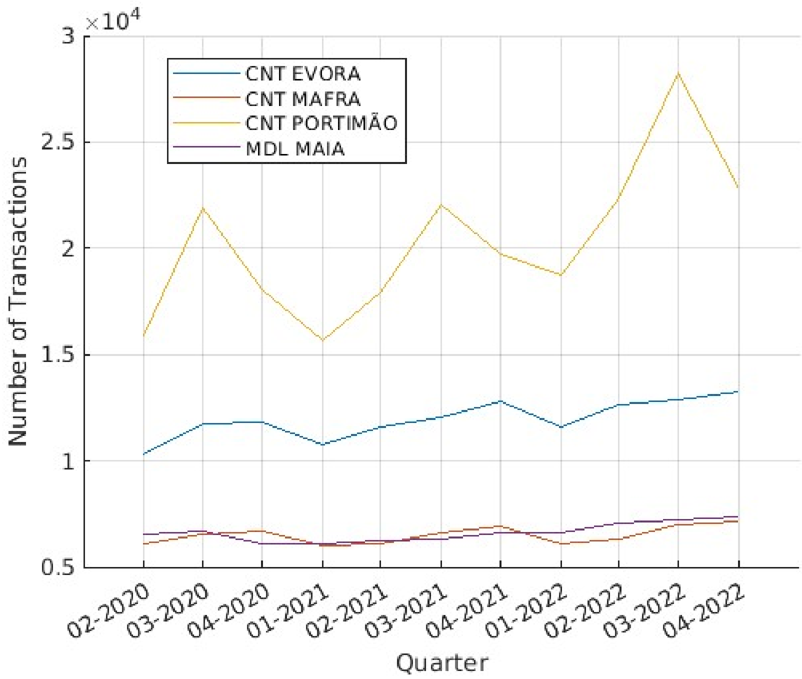
Adding $s \cdot (\eta^N)'$ would be necessary to restore the intensive margin, typically lowering the Lerner index relative to the uncorrected approach.

In this section, I discuss the matter of participation rates and verify the robustness of our estimates of price elasticity of demand both to inter-supermarket competition and price endogeneity.

B.2.1 Transaction trends

Consider the time trends for the number of transactions observed per supermarket in the sample period:

Figure B.1: Per-supermarket transaction trends



The trends do not resemble those observed for revenue-weighted mean prices in Figure 2.2. Nonetheless, the lack of a clear relationship between the data on grocery stores' prices and transactions numbers may mask a participation elasticity. Core inflation rose substantially in Portugal through the period of analysis. Nominal inflation pushed all prices up, but if as a result people shifted spending from elsewhere to groceries, this may have led to more transactions in-store. For example, the price of home food relative to restaurant meals may have fallen, attracting shoppers despite higher nominal grocery prices. This could be masking the correct price elasticities of grocery store purchases.

B.2.2 Estimation of participation rate elasticity

We wish to identify the causal effect of grocery store prices on participation N, net of common inflation and fixed market traits. The proposed estimation strategy is a first-difference IV panel estimator:

$$\Delta \ln(N_{m,t}) = \delta \cdot \Delta \ln P_{m,t} + \tau_t + \Delta u_{m,t} \tag{B.7}$$

Parameter δ is the participation elasticity. The number of transactions is calculated on a per-quarter, per-supermarket basis, while the price index $P_{m,t}$ is set under two specifications: (i) a simple average across goods; and (ii) a revenue-weighted function of per-quarter, per-supermarket goods prices. This approach matches our earlier empirical specification, which used quarterly data, but also limits the size of the regression we can run. Variable τ_t is a time trend.

Taking first differences removes latent time-invariant terms while preempting trend drifts. An instrument is necessary in this setting, as before, to isolate the price effect on participation rates with exogenous variation, avoiding reverse causality. I use the change in a good's competitors' average price (our previous instrument), aggregated into an index as that which it instruments for.

For robustness, I consider current, 1-2 lags and select that with the highest first-stage F. With one endogenous regressor, we run 2SLS and report Eicker-Huber-White standard errors; the 2SLS p-value; the Anderson-Rubin p-value, which stays valid even when instruments are weak; and the (homoskedastic) first-stage F for the chosen instrument set.

Results are shown below:

Table B.1: Participation shares - IV results by specification

Price index	$\hat{\delta}$ (SE)	p-value	AR p-value	First-stage F	N
Simple average	0.056 (2.685)	0.983	0.981	6.920	32
Revenue-weighted	−2.855 (4.890)	0.559	0.223	6.119	32

Notes: $\hat{\delta}$ is the IV estimate; standard errors in parentheses. "AR p-value" is the Anderson-Rubin test p-value. "First-stage F" is the first-stage F-statistic. N is the number of observations.

The preferred instrument is the first lag of the competitor price change with a first-stage $F \approx 6.12$ and 6.92 respectively, still however indicating weak instrument strength. Nonetheless, neither specification shows a detectable effect of prices on participation, even under the Anderson-Rubin test. The simple-average specification yields $\hat{\delta} \approx 0.06$ and the revenue-weighted specification $\hat{\delta} \approx -2.86$. Deflating the price indices by the CPI leaves the results unchanged.

Though underpowered, the results suggest participation sensitivity is not a significant driver of our results. This seems to confirm that (i) inter-supermarket competition is not a major problem in the set of stores and period I have sampled, and (ii) adjusting demand by the number of transactions observed each period in each store raises no endogeneity concerns within our sample. Nonetheless, the approach used here could be averaging out what are likely important effects amongst staple goods in any supermarket assortment. The reader is therefore encouraged to perceive this thesis's mark-up estimates as upper bounds, especially so for goods on the upper tail of the revenue distribution.

C

Appendix - Chapter 3

Contents

C.1 Proofs **147**

C.2 Finding S when X is not observed but M is **150**

C.3 Connection between M and $\mathbb{E}(\boldsymbol{\delta}\boldsymbol{\delta}')$ **151**

C.4 Exploratory analysis of empirical testability **151**

C.4.1 Estimating U 152

C.4.2 Estimating Γ 156

C.1 Proofs

Lemma 3.1. Note that $\text{span}(X) = \text{span}(Z)$. Let $Q = [Z\ Z_\perp]$, an orthonormal basis composed of Z and its orthogonal complement $Z_\perp$. Orthogonal complements span the vector space outside the range of a given matrix; i.e. $Z_\perp$ spans the set of all vectors that are orthogonal to every vector in Z. For M a full-rank matrix (necessary and sufficient for a well-behaved utility function), our assertion that it is partially a function of Z implies that the remaining variation is represented by coordinates in $Z_\perp$.

Let $\hat{M} = E(M|X)$ and:

$$\tilde{M} = Q\hat{M}Q' \tag{C.1}$$

The $\tilde{M}$ matrix is $\hat{M}$ but expressed in the Q-basis. Being symmetric, any $\tilde{M}$ can be written as a block matrix with the following structure:

$$\tilde{M} = Q\hat{M}Q' = \begin{bmatrix} D & B \\ B' & C \end{bmatrix} \tag{C.2}$$

with $D \in \mathbb{R}^{K\times K}$, $B = B' \in \mathbb{R}^{K\times N-K}$, and $C \in \mathbb{R}^{(N-K)\times(N-K)}$. This follows from the fact that $\tilde{M}$ maintains the symmetry properties of $\hat{M}$.

Now, without any information outside the span of Z, i.e. in span$(Z_\perp)$, we may wish to remain agnostic about what is an appropriate approximation for that portion of M. To formalise that ignorance, we may require that $\tilde{M}$ be invariant to orthogonal transformations acting only on the complement subspace. In other words, we adjust $\tilde{M}$ so that it is isotropic outside the span of Z. Isotropy can be defined as follows:

$$\begin{bmatrix} I_K & 0 \\ 0 & H \end{bmatrix} \tilde{M} \begin{bmatrix} I_K & 0 \\ 0 & H \end{bmatrix} = \tilde{M} \tag{C.3}$$

for any matrix $H \in \mathbb{R}^{(N-K)\times(N-K)}$ such that $H'H = HH' = I$. The idea here is that any misspecification in the unobserved component span$(Z_\perp)$ will be less likely to dominate $\hat{M}$ if all possibilities within that component are treated in the same way. Matrix H is used here to show the kind of transformation we want M to be unaffected by - the more affected the expression is by R, the more likely it is that misspecification of $\hat{M}$ in the unobserved component span dominates the matrix, obfuscating the part of M that we do know and are trying to model (i.e. the part dependent on characteristics). From before, we have:

$$\begin{bmatrix} I_K & 0 \\ 0 & H \end{bmatrix} \begin{bmatrix} D & B \\ B' & C \end{bmatrix} \begin{bmatrix} I_K & 0 \\ 0 & H \end{bmatrix} = \begin{bmatrix} D & BH' \\ HB' & HCH' \end{bmatrix} \tag{C.4}$$

we may only have isotropy outside span(Z) if (i) $BH' = B$ and (ii) $HCH' = C$. Take e.g. $H = -I$. Then $BH' = B(-I) = -B$. In other words, a non-zero matrix B is a contradiction: $B = 0$. It is similarly straightforward to show that $C = \rho I$ by first demonstrating that it must be diagonal (or else a reflection matrix can flip the sign of any off-diagonal coordinate and break the equality) and all its diagonal elements must be identical (or else a permutation matrix can change the order of the rows/columns and break the equality).

With $B = 0$ and $C = \rho I$, restate $\tilde{M} = Q\hat{M}Q'$ in terms of $\hat{M}$:

$$\hat{M} = Q'\tilde{M}Q = Q' \begin{bmatrix} D & 0 \\ 0 & \rho I \end{bmatrix} Q = ZDZ' + \rho(I - ZZ') \tag{C.5}$$

This completes the proof. □

Lemma 3.2. There is always a $\boldsymbol{b}$ such that $Z\tilde{\boldsymbol{\beta}} = ZU\boldsymbol{b} \Leftrightarrow Z(\tilde{\boldsymbol{\beta}} - U\boldsymbol{b}) = 0$ if U is invertible; then $\boldsymbol{b} = U^{-1}\tilde{\boldsymbol{\beta}}$. This will be the case if A, from which U is derived, is symmetric, which it is by construction as M is positive definite. Then U, as an orthogonal matrix of A's eigenvectors, is invertible, and we may re-write $\boldsymbol{\delta} = S\boldsymbol{b} + \boldsymbol{v}$, for $S = ZU$. If $\boldsymbol{\beta} > 0$, note that we can always define a $\boldsymbol{b} > 0$, as eigenvectors are defined up to a sign - we can change the signs of the columns of U as needed to achieve this. □

Lemma 3.3. In general, when writing $\mathbb{E}\Pi$, all terms related to $\boldsymbol{v}$ can be ignored. We can see this from $\mathbb{E}\Pi = (-1/(4\phi))\boldsymbol{\delta}'M^{-1}\boldsymbol{\delta} = (-1/(4\phi))(X\boldsymbol{\beta})'M^{-1}(X\boldsymbol{\beta}) + 2(-1/(4\phi))(X\boldsymbol{\beta})'M^{-1}\boldsymbol{v} + (-1/(4\phi))\boldsymbol{v}'M^{-1}\boldsymbol{v}$. The term in the middle drops out as $\mathbb{E}\boldsymbol{v} = 0$. The second term is constant in S: $\mathbb{E}(\boldsymbol{v}'M^{-1}\boldsymbol{v}|X) = \mathbb{E}(\mathrm{tr}(\boldsymbol{v}'M^{-1}\boldsymbol{v}))|X) = \mathbb{E}(\mathrm{tr}(M^{-1}\boldsymbol{v}'\boldsymbol{v})|X) = \mathrm{tr}(M^{-1}\mathbb{E}(\boldsymbol{v}\boldsymbol{v}')|X)$; our unit variance assumption (we may allow this up to a constant, only homoskedasticity is required) together with $\mathbb{E}\boldsymbol{v} = 0$ implies $\mathrm{tr}(M^{-1}\mathbb{E}(\boldsymbol{v}\boldsymbol{v}')|X) = tr(M^{-1}) = \sum_{k=1}^{K} 1/\gamma_k$, which is independent from S. $\square$

C.2 Finding S when X is not observed but M is

Just as we can produce a positive definite matrix approximately from assuming it is partially a function of X using X's orthogonal representation, we can take advantage of the spectral decomposition of a known measure of competition to produce the relevant attributes which drive said competition. Note that, for any matrix, we can obtain the following spectral decomposition:

$$M = \sum_{i=1}^{N} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i', \ \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N \geq 0$$

For an $N \times N$ matrix Σ, we can write:

$$\begin{aligned} M &= \sum_{n=1}^{N} \lambda_n \boldsymbol{u}_n \boldsymbol{u}_i' \\ &= \sum_{n=1}^{N} (\lambda_n - \rho) \boldsymbol{u}_n \boldsymbol{u}_n' + \sum_{i=1}^{N} \rho \cdot \boldsymbol{u}_i \boldsymbol{u}_i' \\ &= \sum_{n=1}^{N} \gamma_n \boldsymbol{u}_n \boldsymbol{u}_n + \rho I_N \end{aligned}$$

for $\gamma_n = \lambda_n - \rho, \ \forall n = 1, \ldots, N$. The $\boldsymbol{u}$ here effectively operate as $\boldsymbol{s}$ in the earlier specification. To keep the assumption of $\gamma_n > 0, \ \forall n$, however, the above equation will only be approximately true; if $\exists \lambda_k : \lambda_k < \rho$, then we must effectively set $\lambda_k = \rho$, as otherwise M will not be positive definite as required. This will nonetheless be at least as good as in the case where we have X but not M, where some of the attributes for which $\gamma > 0$ may not be observed.

C.3 Connection between M and $\mathbb{E}(\boldsymbol{\delta\delta'})$

Our $\hat{M}$ measure can be shown to be proportional to the second (raw) moment of $\boldsymbol{\delta}$, $E(\boldsymbol{\delta\delta'})$. Under the assumptions we have placed earlier on each term in $\boldsymbol{\delta}$:

$$\mathbb{E}(\boldsymbol{\delta\delta'}) = \mathbb{E}(S\boldsymbol{bb'}S') + \mathbb{E}(\boldsymbol{vv'}) = \sum_{k=1}^{K} \gamma_k \boldsymbol{s}_k \boldsymbol{s}'_k + I_N$$

for $\gamma_k = b_k^2$. In this sense, γ_k, $\forall k$ can be framed within a representative consumer setup as a measure of dispersion in mean utility from attribute k along the consumer population. In general, the Chapter assumes $\gamma_k \neq b_k^2$.

It is well-understood in the Statistics literature on common factor models - from which this Chapter draws heavily - that any vector which can be described as a linear function of factors (such as $\boldsymbol{\delta}$ and S above) can, under certain properties, allow the variance-covariance matrix of said factors be described as above (see e.g. Darton, 1980). It is also well-known in the Finance literature that quadratic utility is consistent with mean–variance analysis (see e.g. Ross, 1977; Fama and French, 1993), with the Hessian matrix M encoding expected asset return variance. The above results from the interaction between these two fields, which I repurpose here for use in empirical microeconomics (and more specifically industrial organisation).

C.4 Exploratory analysis of empirical testability

In Chapter 2, an approach was introduced for estimating linear demand. In Chapter 3, we have used the same model with some adjustments. In this Appendix, I explore how such adjustments impact the estimation of the Chapter 2 model. More specifically, I make a first foray into the matter of estimating the parameters exclusive to this Chapter, verifying whether a minimal condition for estimation - the well-behavedness of the model, in respect of these parameters, within the context of linear regression - is satisfied.

C.4.1 Estimating U

Throughout the Chapter, we have worked with $S = ZU$. However, while Z may in general be thought of as observable, through QR decomposition on the matrix of observed characteristics X, U is not. In this section, I break down how we may parameterise U such that it is estimable via econometric methods. I propose the following specification for $U(\boldsymbol{\theta})$:

$$U(\boldsymbol{\theta}) = \prod_{0 \leq i < j \leq K} G(i, j, \theta_{ij}) \tag{C.6}$$

for $G(i, j, \theta_{ij})$ a Givens rotation (Golub and van Loan, 1983, p.240):

$$G(i, j, \theta_{ij}) = \begin{bmatrix} 1 & \cdots & 0 & \cdots & \cdots & \cdots & 0 & \cdots & 0 \\ \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & \cdots & \cos(\theta_{ij}) & 0 & \cdots & \cdots & \sin(\theta_{ij}) & \cdots & 0 \\ \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & \cdots & -\sin(\theta_{ij}) & 0 & \cdots & \cdots & \cos(\theta_{ij}) & \cdots & 0 \\ \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & \cdots & \cdots & \cdots & 0 & \cdots & 1 \end{bmatrix} \tag{C.7}$$

with non-$\{0, 1\}$ elements reflecting where the rows i and j (self-)intersect. A Givens rotation turns two coordinates and leaves all the others unchanged. For example, for a 2-by-2 $U(\boldsymbol{\theta})$, if $\boldsymbol{y} = U(\boldsymbol{\theta})'\boldsymbol{x}$, the $\boldsymbol{y}$ is obtained by rotating $\boldsymbol{x}$ counter clockwise through an angle θ_{ij}.

We can confirm $G(i, j, \theta_{ij})$ is orthogonal: $G'G = I$; $\det(G) = 1$; and $G(i, j, \theta_{ij})^{-1} = G(i, j, -\theta_{ij})$. Furthermore, Givens rotations' flexibility allows us to, by chaining a small number of such two–coordinate rotations, get a rotation along multiple coordinates at once. Each factor uses a chosen coordinate pair (i, j) and one angle θ_{ij}.

If $Z = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K]$ has orthonormal columns (a basis for the span of X), we can turn those directions by $U(\boldsymbol{\theta})$:

$$S = ZU(\boldsymbol{\theta}) = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_K] \tag{C.8}$$

and then proceed as we have so far. Because $U(\boldsymbol{\theta})$ is orthogonal, the $\boldsymbol{s}_k$ remain orthonormal.

Before moving on to studying the well-behaveness of a regression equation relative to the $\boldsymbol{\theta}$ parameters, note the following. Firstly, the order of the multiplication in $U(\boldsymbol{\theta})$ matters: $G(i, k, \theta_{ik})G(i, j, \theta_{ij}) \neq G(i, j, \theta_{ij})G(i, k, \theta_{ik})$, $\forall i, j, k$, so they do not commute. For empirical estimation of our model, however, this does not matter. The $\boldsymbol{\theta}$ parameters are meant only to provide the necessary fit for how our product characteristics X enter M. Secondly, the total number of parameters, for K the number of product attributes, is up to $\frac{K(K-1)}{2}$. This depends only on the number of attributes, not of goods, and therefore likely scales well where $K < N$.

Let us now consider whether, for $U(\boldsymbol{\theta})$ defined via Givens rotations, our regression equation is well-behaved in the $\boldsymbol{\theta}$ parameters. Let:

$$M(\boldsymbol{\theta}) = I + \sum_{k=1}^{K} \gamma_k \boldsymbol{s}_k(\boldsymbol{\theta}) \boldsymbol{s}_k(\boldsymbol{\theta})' \tag{C.9}$$

Then:

$$M^{-1}(\boldsymbol{\theta}) = I - \sum_{k=1}^{K} \alpha_k \boldsymbol{s}_k(\boldsymbol{\theta})\boldsymbol{s}_k(\boldsymbol{\theta})' \tag{C.10}$$

via orthonormality in the columns of S and the Sherman-Woodbury formula, with $\alpha_k = \frac{\gamma_k}{1+\gamma_k} \in [0, 1)$ and $\gamma_p \neq \gamma_q,\ \forall p, q$. I will prove well-behavedness in $M^{-1}(\boldsymbol{\theta})$ without loss of generality, allowing researchers to choose whichever alternative works best for their empirical specification.

Consider the behaviour of the model with regards to parameter θ, which changes the angle between the pair of attribute vectors $\boldsymbol{s}_p(\theta)$ and $\boldsymbol{s}_q(\theta)$). We can write how these vectors depend on θ as follows:

$$[\boldsymbol{s}_p(\theta)\ \boldsymbol{s}_q(\theta)] = [\boldsymbol{s}_p\ \boldsymbol{s}_q]R(\theta) \qquad R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{C.11}$$

The matrix $R(\theta)$ is known as a *rotation matrix.* All other $\boldsymbol{s}_k$ are fixed. Let $\boldsymbol{y}, \boldsymbol{x} \in \mathbb{R}^N$ be given, for $\boldsymbol{y} = M^{-1}\boldsymbol{x}$ a stylised version of the linear regression model of Chapter 2. Decompose without loss of generality $\boldsymbol{y} = \boldsymbol{y}_\perp + y_p\boldsymbol{s}_p + y_q\boldsymbol{s}_q$, $\boldsymbol{x} = \boldsymbol{x}_\perp + x_p\boldsymbol{s}_p + x_q\boldsymbol{s}_q$.[1] Define $\beta_k = 1 - \alpha_k = \frac{1}{1+\gamma_k} > 0$, $\beta_s = \frac{\beta_p+\beta_q}{2}$, and $\beta_d = \frac{\beta_p-\beta_q}{2}$.

Note then that:

$$\begin{aligned} M(\theta)^{-1}\boldsymbol{x}_\perp &= (I - \sum_{k=1}^{K} \alpha_k \boldsymbol{s}_k \boldsymbol{s}_k')\boldsymbol{x}_\perp = \boldsymbol{x}_\perp \\ M(\theta)^{-1}\boldsymbol{s}_j &= (I - \sum_{k=1}^{K} \alpha_k \boldsymbol{s}_k \boldsymbol{s}_k')\boldsymbol{s}_j = \boldsymbol{s}_j - \alpha_j \boldsymbol{s}_j = \beta_j \boldsymbol{s}_j, \forall j \neq p, q \end{aligned} \tag{C.12}$$

This shows that (i) $M^{-1}(\theta)$ acts only upon the vectors in span(Z); and (ii) the regression model is unaffected by changes in θ for vectors that are not p and q. In other

[1] Any vector $\boldsymbol{w}$ can be decomposed into a portion within span(S) and another portion in said vector space's orthogonal complement: i.e. $\boldsymbol{w} = \sum_k w_k \boldsymbol{s}_k + \boldsymbol{w}_\perp$ with $w_k = \boldsymbol{s}_k'\boldsymbol{w}$ and $\boldsymbol{w}_\perp \perp \text{span}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_K\}$, for any $\boldsymbol{w} \in R^N$.

words, we are able to isolate our focus on how the angle between $\boldsymbol{s}_p$ and $\boldsymbol{s}_q$ changes how these variables drive demand.

Let us then isolate the part of M^{-1} which depends on these two variables and is in fact affected by their angle. A simplified M^{-1} may be written as follows:

$$\begin{aligned} M(\theta)^{-1} &= I - \alpha_p \boldsymbol{s}_p(\theta)\boldsymbol{s}_p(\theta)' - \alpha_q \boldsymbol{s}_q(\theta)\boldsymbol{s}_q(\theta)' \\ &= I - [\boldsymbol{s}_p(\theta)\ \boldsymbol{s}_q(\theta)] \begin{bmatrix} \alpha_p & 0 \\ 0 & \alpha_q \end{bmatrix} [\boldsymbol{s}_p(\theta)\ \boldsymbol{s}_q(\theta)]' \\ &= I - [\boldsymbol{s}_p\ \boldsymbol{s}_q] R(\theta) \begin{bmatrix} \alpha_p & 0 \\ 0 & \alpha_q \end{bmatrix} R(\theta)'[\boldsymbol{s}_p\ \boldsymbol{s}_q]' \end{aligned} \tag{C.13}$$

Writing $M^{-1}\boldsymbol{w}$, for $\boldsymbol{w}$ any vector in the span of vectors $\boldsymbol{s}_p$ and $\boldsymbol{s}_q$ (i.e. $\boldsymbol{w} = [\boldsymbol{s}_p\ \boldsymbol{s}_q]\boldsymbol{u}$, for $\boldsymbol{u} \in \mathbb{R}^2$) and applying the fact that the outcome variable $\boldsymbol{y}$ is only relevant to our purposes within the same vector span ($\boldsymbol{y}$ outside this span are unaffected by θ):

$$\begin{aligned} M^{-1}(\theta)\boldsymbol{w} &= (I - [\boldsymbol{s}_p\ \boldsymbol{s}_q] R(\theta) \begin{bmatrix} \alpha_p & 0 \\ 0 & \alpha_q \end{bmatrix} R(\theta)'[\boldsymbol{s}_p\ \boldsymbol{s}_q]')[\boldsymbol{s}_p\ \boldsymbol{s}_q]\boldsymbol{u} \\ &= [\boldsymbol{s}_p\ \boldsymbol{s}_q](I_2 - R(\theta) \begin{bmatrix} \alpha_p & 0 \\ 0 & \alpha_q \end{bmatrix} R(\theta)')\boldsymbol{u} \\ &= [\boldsymbol{s}_p\ \boldsymbol{s}_q](R(\theta) \begin{bmatrix} \beta_p & 0 \\ 0 & \beta_q \end{bmatrix} R(\theta)')\boldsymbol{u} \\ \Leftrightarrow\ M^{-1}(\theta)\Big|_{pq} &= R(\theta) \begin{bmatrix} \beta_p & 0 \\ 0 & \beta_q \end{bmatrix} R(\theta)' = R(\theta) diag(\beta_p,\ \beta_q) R(\theta)' \end{aligned} \tag{C.14}$$

for $M^{-1}(\theta)\Big|_{pq}$ defining the M^{-1} term restricted within coordinates relative to the fixed orthonormal basis ($\boldsymbol{s}_p\ \boldsymbol{s}_q$). The expression $M^{-1}(\theta)\Big|_{pq}\boldsymbol{x}$ can be expressed as the part of $\boldsymbol{y}$ which is defined over this basis and affected by the change in θ.

Expanding $R(\theta)\text{diag}(\beta_p, \beta_q)R(\theta)'$, for $\cos 2\theta = \cos^2\theta - \sin^2\theta$ and $\sin 2\theta = 2\sin\theta\cos\theta$, we have:

$$
\begin{aligned}
&R(\theta)\mathrm{diag}(\beta_p, \beta_q)R(\theta)' \\
&= \begin{bmatrix} \beta_p \cos^2\theta + \beta_q \sin^2\theta & (\beta_p - \beta_q)\cos\theta\sin\theta \\ (\beta_p - \beta_q)\cos\theta\sin\theta & \beta_p \sin^2\theta + \beta_q \cos^2\theta \end{bmatrix} \\
&= \beta_s I + \beta_d \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}
\end{aligned}
\tag{C.15}
$$

such that $\boldsymbol{y} = M(\theta)^{-1}\Big|_{pq}\boldsymbol{x}$ is equivalent to:

$$
\boldsymbol{y} = \beta_s \boldsymbol{x} + \beta_d \begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix} \boldsymbol{x}
\tag{C.16}
$$

This does two things: it isolates the part of $\boldsymbol{y}$ which is affected by θ, and, given $\boldsymbol{y}$, $\boldsymbol{x}$ are observed and β_s is fixed, $\boldsymbol{r} = \boldsymbol{y} - \beta_s\boldsymbol{x}$ can be defined, allowing an avenue to estimating θ. To see how, notice the following must hold:

$$
\begin{aligned}
&\beta_d \begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix} \boldsymbol{x} = \beta_d \begin{bmatrix} x_p & x_q \\ -x_q & x_p \end{bmatrix} \begin{bmatrix} \cos(2\theta) \\ \sin(2\theta) \end{bmatrix} = \boldsymbol{r} \\
\Leftrightarrow &\begin{bmatrix} \cos(2\theta) \\ \sin(2\theta) \end{bmatrix} = \frac{1}{\beta_d ||\boldsymbol{x}||^2} \begin{bmatrix} x_p & -x_q \\ x_q & x_p \end{bmatrix} \boldsymbol{r}
\end{aligned}
\tag{C.17}
$$

assuming $\beta_d||\boldsymbol{x}||^2 \neq 0$; and while a unique solution, any LHS satisfying this equation only reflects an optimal angle θ if $||(\cos(2\theta)\ \ \sin(2\theta))'|| = 1$. We therefore have a unique solution in θ if an only if these conditions apply, though they may be approximated. Then, we can find what that θ is as follows:

$$
\theta = \frac{1}{2}\mathrm{atan2}(\hat{\sin(2\theta)}, \hat{\cos(2\theta)})
\tag{C.18}
$$

unique up to adding π - so any optimisation must have in mind a $\theta \in [0, \pi)$.

C.4.2 Estimating Γ

Estimating Γ is far simpler. Under the orthonormality of S, it follows from (3.22) that:

$$M^{-1} = I - \sum_{k=1}^{K} \frac{\gamma_k}{1+\gamma_k} \boldsymbol{s}_k \boldsymbol{s}_k' \tag{C.19}$$

Defining $\beta_k = \frac{\gamma_k}{1+\gamma_k}$, $\forall k = 1, \ldots, K$, the vector $\boldsymbol{\beta}$ can be estimated in the same way as the $\boldsymbol{\alpha}$ parameters in (2.10). As a matter of fact, insofar as S can be obtained from Z and $U(\boldsymbol{\theta})$ in the previous section, $S'S$ behaves like our proxy W in Chapter 2.