

NBER WORKING PAPER SERIES

MICROECONOMETRIC MODELS OF CONSUMER DEMAND

Jean-Pierre H. Dubé

Working Paper 25215

<http://www.nber.org/papers/w25215>

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue

Cambridge, MA 02138

November 2018

Dubé acknowledges the support of the Kilts Center for Marketing and the Charles E. Merrill faculty research fund for research support. I would like to thank Greg Allenby, Shirsho Biswas, Oeystein Daljord, Stefan Hoderlein, Joonhwi Joo, Kyeongbae Kim, Yewon Kim, Nitin Mehta, Olivia Natan, and Robert Sanders for helpful comments and suggestions. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2018 by Jean-Pierre H. Dubé. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Microeconomic Models of Consumer Demand
Jean-Pierre H. Dubé
NBER Working Paper No. 25215
November 2018
JEL No. D11,D12,L66,M3

ABSTRACT

A long literature has developed econometric methods for estimating individual-consumer-level demand systems that accommodate corner solutions. The increasing access to transaction-level customer purchase histories across a wide array of markets and industries vastly expands the prospect for improved customer insight, more targeted marketing policies and individualized welfare analysis. A descriptive analysis of a broad, CPG database indicates that most consumer brand categories offer a wide variety of differentiated offerings for consumers. However, consumers typically purchase only a limited scope of the available variety, leading to a very high incidence of corner solutions which poses computational challenges for demand modeling. Historically, these computational challenges have limited the applicability of microeconomic models of demand in practice, except for the special case of pure discrete choice (e.g., logit and probit). Recent advances in computing power along with methods for numerical and simulation-based integration have been instrumental in facilitating the broader use of these models in practice. We survey herein the extant literature on the neoclassical derivation of microeconomic demand models that allow for corner solutions and differentiated products. We summarize the key developments in the literature, including the role of consumers' price expectations, and point towards opportunities for future research.

Jean-Pierre H. Dubé
University of Chicago
Booth School of Business
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
jdube@chicagobooth.edu

Contents

1	Introduction	3
2	Empirical Regularities in Shopping Behavior: the CPG Laboratory	7
3	The Neoclassical Derivation of an Empirical Model of Individual Consumer Demand	10
3.1	The Neoclassical Model of Demand with Binding, Non-Negativity Constraints . . .	11
3.1.1	Estimation Challenges with the Neoclassical Model	14
3.1.2	Example: Quadratic Utility	16
3.1.3	Example: Linear Expenditure System (LES)	18
3.1.4	Example: Translated CES Utility	19
3.1.5	Virtual Prices and the Dual Approach	22
3.1.6	Example: Indirect Translog Utility	25
3.2	The Discrete/Continuous Product Choice Restriction in the Neoclassical Model . .	26
3.2.1	The Primal Problem	26
3.2.2	Example: Translated CES Utility	28
3.2.3	Example: the Dual Problem with Indirect Translog Utility	29
3.2.4	Promotion Response: empirical findings using the discrete/continuous de- mand model	31
3.3	Indivisibility and the Pure Discrete Choice Restriction in the Neoclassical Model .	32
3.3.1	A Neoclassical Derivation of the Pure Discrete Choice Model of Demand .	32
3.3.2	The Classic Pure Discrete Choice Model of Demand	35
4	Some Extensions to the Typical Neoclassical Specifications	39
4.1	Income Effects	39
4.1.1	A Non-Homothetic Discrete Choice Model	41
4.2	Complementary Goods	43
4.2.1	Complementarity Between Products Within a Commodity Group	45

4.2.2	Complementarity Between Commodity Groups (multi-category models) . .	48
4.3	Discrete Package Sizes and Non-Linear Pricing	52
4.3.1	Expand the Choice Set	53
4.3.2	Models of Pack Size Choice	55
5	Moving Beyond the Basic Neoclassical Framework	56
5.1	Stock-Piling, Purchase Incidence and Dynamic Behavior	56
5.1.1	Stock-Piling and Exogenous Consumption	57
5.1.2	Stock-Piling and Endogenous Consumption	60
5.1.3	Empirical Findings with Stock-piling Models	63
5.2	The Endogeneity of Marketing Variables	64
5.2.1	Incorporating the Supply Side: A Structural Approach	68
5.2.2	Incorporating the Supply Side: A Reduced-Form Approach	70
5.3	Behavioral Economics	73
5.3.1	The Fungibility of Income	73
5.3.2	Social Preferences	75
6	Conclusions	79

1 Introduction

A long literature in quantitative marketing has used the structural form of microeconomic models of demand to analyze consumer-level purchase data and conduct inference on consumer behavior. These models have played a central role in the study of some of the key marketing questions, including the measurement of brand preferences and consumer tastes for variety, the quantification of promotional response, the analysis of the launch of new products and the design of targeted marketing strategies. In sum, the structural form of the model is critical for measuring unobserved economic aspects of consumer preferences and for simulating counter-factual marketing policies that are not observed in the data (e.g. demand for a new product that has yet to be launched).

The empirical analysis of consumer behavior is perhaps one of the key areas of overlap between the economics and marketing literatures. The application of empirical models of aggregate demand using restrictions from microeconomics dates back at least since the mid 20th century (e.g., Stone, 1954). Demand estimation plays a central role in marketing decision-making. Marketing-mix models, or models of demand that account for the causal effects of marketing decision variables, such as price, promotions and other marketing tools, are fundamental for the quantification of different marketing decisions. Examples include the measurement of market power, the measurement of sales-reponse to advertising, the analysis of new product introductions and the measurement of consumer welfare, just as a few examples.

Historically, the data used for demand estimation typically consisted of market-level, aggregate sales quantities under different marketing conditions. In the digital age, the access to transaction-level data at the point of sale has become nearly ubiquitous. In many settings, firms can now assemble detailed, longitudinal databases tracking individual customers' purchase behavior over time and across channels. Simply aggregating these data for the purposes of applying traditional aggregate demand estimation techniques creates several problems. First, aggregation destroys potentially valuable information about customer behavior. Besides the loss of potential statistical

efficiency, aggregation eliminates the potential for individualized demand analysis. Innovative selling technologies have facilitated a more segmented and even individualized approach to marketing, requiring a more intimate understanding of the differences in demand behavior between customer segments or even between individuals. Second, aggregating individual demands across customers facing heterogeneous marketing conditions can create biases that could have adverse effects on marketing decision-making (e.g., Gupta, Chintagunta, Kaul, and Wittink, 1996)¹.

Our discussion herein focuses on microeconomic models of demand designed for the analysis of individual consumer-level data. The microeconomic foundations of a demand model allow the analyst to assign a structural interpretation to the model's parameters, which can be beneficial for assessing "consumer value creation" and for conducting counter-factual analyses. In addition, as we discuss herein, the cross-equation restrictions derived from consumer theory can facilitate more parsimonious empirical specifications of demand. Finally, the structural foundation of the econometric uncertainty as a model primitive provides a direct correspondence between the likelihood function and the underlying microeconomic theory.

Some of the earliest applications of microeconomic models to marketing data analyzed the decomposition of consumer responses to temporary promotions at the point of purchase (e.g., Guadagni and Little, 1983; Chiang, 1991; Chintagunta, 1993). Of particular interest was the relative extent to which temporary price discounts caused consumers to switch brands, increase consumption, or strategically forward-buy to stockpile during periods of low prices. The microeconomic approach provided a parsimonious, integrated framework to with which understand the inter-relationship between these decisions and consumer preferences. At the same time, the cross-equation restrictions from consumer theory can reduce the degrees-of-freedom in an underlying statistical model used to predict these various components of demand.

Most of the foundational work derives from the consumption literature (see Deaton and Muellbauer, 1980b, for an extensive overview). The consumption literature often emphasizes the use of cost functions and duality concepts to simplify the implementation of the restrictions from eco-

¹Blundell, Pashardes, and Weber (1993) find that models fit to aggregate data generate systematic biases relative to models fit to household-level data, especially in the measurement of income effects.

conomic theory. In this survey, we mostly focus on the more familiar direct utility maximization problem. The use of a parametric utility function facilitates the application of demand estimates to broader topics than the analysis of price and income effects, such as product quality choice, consumption indivisibilities, product positioning, and product design.

In addition, our discussion focuses on a very granular, product-level analysis within a product category². Unlike the macro-consumption literature, which focuses on budget allocations across broad commodity groups like food, leisure, and transportation, we focus on consumer's brand choices within a narrow commodity group, such as the brand variants and quantities of specific laundry detergents or breakfast cereals purchased on a given shopping trip. The role of brands and branding have been shown to be central to the formation of industrial market structure (e.g., Bronnenberg, Dhar, and Dubé, 2005; Bronnenberg and Dubé, 2017). To highlight some of the differences between a broad commodity-group focus versus a granular brand-level focus, we begin the chapter with a short descriptive exercise laying out several key stylized facts for households' shopping behavior in consumer packaged goods (hereafter CPG) product categories using the Nielsen-Kilts Homescan database. We find that the typical consumer goods category offers a wide array of differentiated product alternatives available for sale to consumers, often at different prices and under different marketing conditions. Therefore, consumer behavior involves a complex trade-off between the prices of different goods and their respective perceived qualities. Moreover, an individual household typically purchases only a limited scope of the variety available. This purchase behavior leads to the well-known "corner solutions" problem whereby expenditure on most goods is typically zero. Therefore, a satisfactory microeconomic model needs to be able to accommodate a demand system over a wide array of differentiated offerings and a high incidence of corner solutions.

The remainder of the chapter surveys the neoclassical, microeconomic foundations of models of individual demand that allow for corner solutions³. From an econometric perspective, non-

²We refer readers interested in a discussion of aggregation and the derivation of models designed for estimation with market-level data to the surveys by Deaton and Muellbauer (1980b), Nevo (2011) and Pakes (2014).

³See also the following for surveys of microeconomic models: Nair and Chintagunta (2011) for marketing, Phaneuf and Smith (2005) for environmental economics, and Deaton and Muellbauer (1980b) for the consumption

purchase behavior contains valuable information about consumers' preferences and the application of econometric models that impose strictly interior solutions would likely produce biased and inconsistent estimates of demand - a selection bias. However, models with corner solutions introduce a number of complicated computational challenges, including high-dimensional integration over truncated distributions and the evaluation of potentially complicated Jacobian matrices. The challenges associated with corner solutions have been recognized at least since Houthaker (1953) and Houthakker (1961) who discuss them as a special case of quantity rationing. We also discuss the role of discreteness both in the brand variants and quantities purchased. In particular, we explore the relationship between the popular discrete choice models of demand (e.g., logit) and the more general neoclassical models.

In a follow-up section, we discuss several important extensions of the empirical specifications used in practice. We discuss the role of income effects. For analytic tractability, many popular specifications impose homotheticity and quasi-linearity conditions that limit or eliminate income effects. We discuss non-homothetic versions of the classic discrete choice models that allow for more realistic asymmetric substitution patterns between vertically-differentiated goods.

Another common restriction used in the literature is additive separability both across commodity groups and across the specific products available within a commodity group. This additivity implies that all products are gross substitutes, eliminating any scope for complementarity across goods. We discuss recent research that has analyzed settings with complementary goods.

In many consumer goods categories, firms use complex non-linear pricing strategies that restrict the quantities a consumer can purchase to a small set of pre-packaged commodity bundles. We do not discuss the price discrimination itself, focusing instead on the indivisibility the commodity bundling imposes on demand behavior.

In the final section of the survey, we discuss several important departures from the standard neoclassical framework. While most of the literature has focused on static models of brand choices, the timing of purchases can play an important role in understanding the impact of price promotions

literature.

on demand. We discuss dynamic extensions that allow consumers to stock-pile storable goods based on their price expectations. The accommodation of purchase timing can lead to very different inferences about the price elasticity of demand.

We also discuss the potential role of the supply side of the market and the resulting endogeneity biases associated with the strategic manner in which point-of-purchase marketing variables are determined by firms. Most of the literature on microeconomic models of demand has ignored such potential endogeneity in marketing variables.

Finally, we address the emerging area of structural models of behavioral economics that challenge some of the basic elements of the neoclassical framework. We discuss recent evidence of mental accounting in the income effect that creates a surprising non-fungibility across different sources of purchasing power. We also discuss the role of social preferences and models of consumer-response to cause marketing campaigns.

Several important additional extensions are covered in later chapters of this volume, including the role of consumer search, information acquisition and the formation of consideration sets (Chapter 5), the role of brands and branding (Chapter 4), and the role of durable goods and the timing of consumer adoption throughout the product life cycle (Chapter 7). Perhaps the most crucial omission herein is the discussion of taste heterogeneity, which is covered in depth in Chapter 2 of this volume. Consumer heterogeneity plays a central role in the literature on targeted marketing.

2 Empirical Regularities in Shopping Behavior: the CPG Laboratory

In this section, we document broad patterns of purchase behavior across US households in the consumer packaged goods (CPG) industry. We will use these shopping patterns in section 3 as the basis for deriving a microeconomic demand estimation framework derived from neoclassical consumer theory.

The CPG industry represents a valuable laboratory in which to study consumer behavior. CPG

brands are widely available across store formats including grocery stores, supermarkets, discount and club stores, drug stores and convenience stores. They are also purchased at a relatively high frequency. The average US household consumer conducted 1.7 grocery trips per week in 2017⁴. Most importantly, CPG spending represents a sizable portion of household budgets. In 2014, the global CPG sector was valued at \$8 trillion and was predicted to grow to \$14 trillion by 2025⁵. In 2016, US households spent \$407 billion on CPGs⁶. A long literature in brand choice has used household panel data in CPG categories not only due to the economic relevance, but also due to the high quality of the data. CPG categories exhibit high-frequency price promotions that can be exploited for demand estimation purposes.

We use the Nielsen Homescan panel housed by the Kilts Center for Marketing at the University of Chicago Booth School of Business to document CPG purchase patterns. The Homescan panelists are nationally representative⁷. The database tracks purchases in 1,011 CPG product categories (denoted by Nielsen's product module codes) for over 132,000 households between 2004 and 2012, representing over 88 million shopping trips. Nielsen classifies product categories using module codes. Examples of product modules include Carbonated Soft Drinks, Ready-to-Eat Cereals, Laundry Detergents and Tooth Paste.

We retain the 2012 transaction data to document several empirical regularities in shopping behavior. In 2012, we observe 52,093 households making over 6.57 million shopping trips during which they purchase over 46 million products.

The typical CPG category offers a wide amount of variety to consumers. Focusing only on the products actually purchased by Homescan panelists in 2012, the average category offers 402.8 unique products as indexed by a universal product code (UPC) and 64.4 unique brands. For in-

⁴Source: "Consumers' weekly grocery shopping trips in the United States from 2006 to 2017," Statista, 2017, accessed at <https://www.statista.com/statistics/251728/weekly-number-of-us-grocery-shopping-trips-per-household/> on 11/13/2017.

⁵Source: "Three myths about growth in consumer packaged goods," by Rogerio Hirose, Renata Maia, Anne Martinez, and Alexander Thiel, McKinsey, June 2015, accessed at <https://www.mckinsey.com/industries/consumer-packaged-goods/our-insights/three-myths-about-growth-in-consumer-packaged-goods> on 11/13/2017.

⁶source: "Consumer packaged goods (CPG) expenditure of U.S. consumers from 2014 to 2020," Statista, 2017, accessed at <https://www.statista.com/statistics/318087/consumer-packaged-goods-spending-of-us-consumers/> on 11/13/2017

⁷See Einav, Leibtag, and Nevo (e.g., 2010) for a validation study of the Homescan data.

stance, a brand might be any UPC coded product with the brand name Coca Cola, whereas a UPC might be a 6-pack of 12-oz cans of Coca Cola. While the subset of available brands and sizes varies across stores and regions, these numbers reveal the extent of variety available to consumers. In addition, CPG products are sold in pre-packaged, indivisible pack sizes. The average category offers 31.9 different pack size choices. The average brand within a category is sold in 5.4 different pack sizes. Therefore, consumers face an interesting indivisibility constraint, especially if they are determined to buy a specific brand variant. Moreover, CPG firms' widespread pre-commitment to specific sizes is suggestive of extensive use of non-linear pricing.

For the average category, we observe 39,787 trips involving at least one purchase. Households purchase a single brand and a single pack during 94.3% and 67.3% of the category-trip combinations, respectively. On average, households purchase 1.07 brands per category-trip. In sum, the discrete brand choice assumption, and to a lesser extent the discrete quantity choice assumption, seems broadly appropriate across trips at the category level.

However, we do observe categories in which the contemporaneous purchase of assortments is more commonplace and many of these categories are economically large. In the Ready-to-Eat Cereals category, which ranks third overall in total household expenditures among all CPG categories, consumers purchase a single brand during only 72.6% of trips. Similarly, for Carbonated Soft Drinks and Refrigerated Yogurt, which rank fourth and tenth overall respectively, consumers purchase a single brand during only 81.5% and 86.6% of trips respectively. Therefore, case studies of some of the largest CPG categories may need to consider demand models that allow for the purchase of variety, even though only a small number of the variants is chosen on any given trip. Similarly, we observe many categories where consumers occasionally purchase multiple packs of a product, even when only a single brand is chosen. In these cases, a demand model that accounts for the intensive margin of quantity purchased may be necessary. We also observe brand switching across time within a category, especially in some of the larger categories. For instance, during the course of the year, households purchased 7.5 brands of Ready-to-Eat Cereals (ranked 3rd), 5.9 brands of Cookies (ranked 11th), 4.7 brands of Bakery Bread (ranked 7th), and 4.6 brands of

Carbonated Soft Drinks (ranked 4th). In many of the categories with more than an average of 3 brands purchased per household-year, we typically observe only one brand being chosen during an individual trip.

In summary, a snapshot of a single year of CPG shopping behavior by a representative sample of consumers indicates some striking patterns. In spite of the availability of a large amount of variety, on any given trip, consumers purchase only a very small number of variants. From a modeling perspective, we observe a high incidence of corner solutions. In some of the largest product categories, consumers routinely purchase assortments, leading to complex patterns of corner solutions. In most categories, the corner solutions degenerate to a pure discrete choice scenario where a single unit of a single product is purchased. In these cases, the standard discrete choice models may be sufficient. However, the single unit is typically one of several pre-determined pack sizes available suggesting an important role for indivisibility on the demand side, and non-linear pricing on the supply side.

3 The Neoclassical Derivation of an Empirical Model of Individual Consumer Demand

The empirical regularities in section 2 show that household-level demand for consumer goods exhibit a high incidence of corner solutions: purchase occasions with zero expenditure on most items in the choice set. The methods developed in the traditional literature on demand estimation (e.g., Deaton and Muellbauer, 1980b) do not accommodate zero consumption. In this section, we review the formulation of the neoclassical consumer demand problem and the corresponding challenges with the accommodation of corner solutions into an empirical framework. Our theoretical starting point is the usual static model of utility maximization whereby the consumer spends a fixed budget on a set of competing goods. Utility theory plays a particularly crucial role in accommodating the empirical prominence of corner solutions in individual-level data.

3.1 The Neoclassical Model of Demand with Binding, Non-Negativity Constraints

We start with the premise that the analyst has access to marketing data comprising individual-level transactions. The analyst's data include the exact vector of quantities purchased by a customer on a given shopping trip, $\hat{\mathbf{x}} = (\hat{x}_1, \dots, \hat{x}_{J+1})'$. An individual transaction database typically has a panel format with time-series observations (trips) for a cross section of customers. We assume that the point-of-purchase causal environment consists of prices, but the database could also include other marketing promotional variables. Our objective consists of deriving a likelihood for this observed vector of purchases from microeconomic primitives. Suppose WLOG that the consumer does not consume the first l goods: $\hat{x}_j = 0$ ($j = 1, \dots, l$), and $\hat{x}_j > 0$ ($j = l + 1, \dots, J + 1$).

We use the neoclassical approach to deriving consumer demand from the assumption that each consumer maximizes a utility function $U(\mathbf{x}; \theta, \varepsilon)$ defined over the quantity of goods consumed, $\mathbf{x} = (x_1, \dots, x_{J+1})'$. Since most marketing studies focus on demand behavior within a specific "product category," we adopt the terminology of Becker (1965) and distinguish between the "commodity" (e.g., the consumption benefit of the category, such as laundry detergent), and the $j = 1, \dots, J$ "market goods" to which we will refer as "products" (e.g., the various brands sold within the product category, such as Tide and Wisk laundry detergents). The quantities in \mathbf{x} are non-negative ($x_j \geq 0 \forall j$) and satisfy the consumer's budget constraint $\mathbf{x}'\mathbf{p} \leq y$, where $\mathbf{p} = (p_1, \dots, p_{J+1})'$ is a vector of strictly positive prices and y is the consumer's budget. The vector θ consists of unknown (to the researcher) parameters describing the consumer's underlying preferences and the vector ε captures unobserved (to the researcher), mean-zero, consumer-specific utility disturbances⁸. Typically ε is assumed to be known to the consumer prior to decision-making.

Formally, the utility maximization problem can be written as follows

$$V(\mathbf{p}, y; \theta, \varepsilon) \equiv \max_{\{x_1, \dots, x_J\}} \{U(\mathbf{x}; \theta, \varepsilon) : \mathbf{x}'\mathbf{p} \leq y, \mathbf{x} \geq 0\} \quad (1)$$

⁸It is straightforward to allow for additional persistent, unobserved taste heterogeneity by indexing the parameters themselves by consumer (see Chapter 2 of this volume).

where we assume $U(\bullet; \theta, \varepsilon)$ is a continuously-differentiable, quasi-concave and increasing function⁹. We can define the corresponding Lagrangian function $L = U(\mathbf{x}; \theta, \varepsilon) + \lambda_y(y - \mathbf{p}'\mathbf{x}) + \lambda_x'\mathbf{x}$ where λ_y and the vector λ_x are Lagrange multipliers for the budget and non-negativity constraints respectively.

A solution to (1) exists as long as the following necessary and sufficient Karush-Kuhn-Tucker (KKT) conditions hold

$$\begin{aligned} \frac{\partial U(\mathbf{x}^*; \theta, \varepsilon)}{\partial x_j} - \lambda_y p_j + \lambda_{x,j} &= 0 & , j = 1, \dots, J+1 \\ y - \mathbf{p}'\mathbf{x}^* &= 0, (y - \mathbf{p}'\mathbf{x}^*)\lambda_y = 0, \lambda_y > 0 & (2) \\ x_j^* \geq 0, x_j^* \lambda_{x,j} &= 0, \lambda_{x,j} \geq 0 & j = 1, \dots, J+1. \end{aligned}$$

Since $U(\bullet; \theta, \varepsilon)$ is increasing, the consumer spends her entire budget (the “adding-up” condition) and at least one good will always be consumed. We define the $J+1$ good as an “essential” numeraire with corresponding price $p_{J+1} = 1$ and with preferences that are separable from those over the commodity group¹⁰. We assume additional regularity conditions on $U(\bullet; \theta, \varepsilon)$ to ensure that an interior quantity of $J+1$ is always consumed: $\frac{\partial U(\mathbf{x}^*; \theta, \varepsilon)}{\partial x_{J+1}} = \lambda_y$ and $\lambda_{x,J+1} = 0$. Therefore, the model can accommodate the case where only the outside good is purchased and none of the inside goods are chosen. We can now re-write the KKT conditions as follows

$$\begin{aligned} \frac{\partial U(\mathbf{x}^*; \theta, \varepsilon)}{\partial x_j} - \frac{\partial U(\mathbf{x}^*; \theta, \varepsilon)}{\partial x_{J+1}} p_j + \lambda_{x,j} &= 0 \quad , j = 1, \dots, J \\ y - \mathbf{p}'\mathbf{x}^* &= 0 & (3) \\ x_j^* \geq 0, x_j^* \lambda_{x,j} &= 0, \lambda_{x,j} \geq 0 & j = 1, \dots, J. \end{aligned}$$

⁹These sufficient conditions ensure the existence of a demand function with a unique consumption level that maximizes utility at a given set of prices (e.g., Mas-Colell, Whinston, and Green, 1995, chapter 3)

¹⁰The essential numeraire is typically interpreted as expenditures outside of the commodity group(s) of interest.

For our observed consumer, recall that $\hat{x}_j = 0$ ($j = 1, \dots, l$) and $\hat{x}_j > 0$ ($j = l + 1, \dots, J + 1$). We can now re-write the KKT conditions to capture these non-consumption decisions

$$\frac{\partial U(\mathbf{x}^*; \theta, \varepsilon)}{\partial x_j} - \frac{\partial U(\mathbf{x}^*; \theta, \varepsilon)}{\partial x_{J+1}} p_j \leq 0 \quad , \quad j = 1, \dots, l \quad (4)$$

$$\frac{\partial U(\mathbf{x}^*; \theta, \varepsilon)}{\partial x_j} - \frac{\partial U(\mathbf{x}^*; \theta, \varepsilon)}{\partial x_{J+1}} p_j = 0 \quad , \quad j = l + 1, \dots, J$$

It is instructive to consider how the KKT conditions (4) influence demand estimation. The $l + 1$ to J equality conditions in (4) implicitly characterize the conditional demand equations for the purchased goods. The l inequality conditions in (4) give rise to the following demand regime-switching conditions, or “selection” conditions

$$\frac{\frac{\partial U(\mathbf{x}^*; \theta, \varepsilon)}{\partial x_j}}{\frac{\partial U(\mathbf{x}^*; \theta, \varepsilon)}{\partial x_{J+1}}} \leq p_j, \quad j = 1, \dots, l \quad (5)$$

which determine whether a given product’s prices are above the consumer’s reservation value, $\frac{\frac{\partial U(\mathbf{x}^*; \theta, \varepsilon)}{\partial x_j}}{\frac{\partial U(\mathbf{x}^*; \theta, \varepsilon)}{\partial x_{J+1}}}$ (see Lee and Pitt, 1986; Ransom, 1987, for a discussion of the switching regression interpretation). We can now see how dropping the observations with zero consumption will likely result in selection bias due to the correlation between the switching probabilities and the utility shocks, ε .

To complete the model, we need to allow for some separability of the utility disturbances. For instance, we can assume an additive, stochastic log-marginal utility: $\ln \left(\frac{\partial U(\mathbf{x}^*; \theta, \varepsilon)}{\partial x_j} \right) = \ln (\bar{U}_j(\mathbf{x}^*; \theta)) + \varepsilon_j$ for each j , where $\bar{U}_j(\mathbf{x}^*; \theta)$ is deterministic. We also assume that ε are random variables with known distribution and density, $F(\varepsilon)$ and $f(\varepsilon)$ respectively. We can now write the KKT conditions more compactly:

$$\tilde{\varepsilon}_j \equiv \varepsilon_j - \varepsilon_{J+1} \leq h_j(x^*; \theta) \quad , \quad j = 1, \dots, l \quad (6)$$

$$\tilde{\varepsilon}_j \equiv \varepsilon_j - \varepsilon_{J+1} = h_j(x^*; \theta) \quad , \quad j = l + 1, \dots, J$$

where $h_j(x^*; \theta) = \ln (\bar{U}_{J+1}(\mathbf{x}^*; \theta)) - \ln (\bar{U}_j(\mathbf{x}^*; \theta)) + \ln (p_j)$.

We can now derive the likelihood function associated with the observed consumption vector, $\hat{\mathbf{x}}$. In the case where all the goods are consumed, then the density of $\hat{\mathbf{x}}$ is

$$f_x(\hat{\mathbf{x}}; \theta) = f_{\tilde{\boldsymbol{\varepsilon}}}(\tilde{\boldsymbol{\varepsilon}}) |J(\hat{\mathbf{x}})| \quad (7)$$

where $J(\hat{\mathbf{x}})$ is the Jacobian of the transformation from $\tilde{\boldsymbol{\varepsilon}}$ to \mathbf{x} . If only the $J + 1$ numeraire good is consumed, the density of $\hat{\mathbf{x}} = (0, \dots, 0)$ is

$$f_x(\hat{\mathbf{x}}; \theta) = \int_{-\infty}^{h_J(\hat{\mathbf{x}}; \theta)} \cdots \int_{-\infty}^{h_1(\hat{\mathbf{x}}; \theta)} f_{\tilde{\boldsymbol{\varepsilon}}}(\tilde{\boldsymbol{\varepsilon}}) d\tilde{\boldsymbol{\varepsilon}}_1 \cdots d\tilde{\boldsymbol{\varepsilon}}_J. \quad (8)$$

For the more general case in which the first l goods are not consumed, the density of $\hat{\mathbf{x}} = (0, \dots, 0, \hat{\mathbf{x}}_{l+1}, \dots, \hat{\mathbf{x}}_J)$ is

$$f_x(\hat{\mathbf{x}}; \theta) = \int_{-\infty}^{h_l(\hat{\mathbf{x}}; \theta)} \cdots \int_{-\infty}^{h_1(\hat{\mathbf{x}}; \theta)} f_{\tilde{\boldsymbol{\varepsilon}}}(\tilde{\boldsymbol{\varepsilon}}_1, \dots, \tilde{\boldsymbol{\varepsilon}}_l, h_{l+1}(\hat{\mathbf{x}}; \theta), \dots, h_J(\hat{\mathbf{x}}; \theta)) |J(\hat{\mathbf{x}})| d\tilde{\boldsymbol{\varepsilon}}_1 \cdots d\tilde{\boldsymbol{\varepsilon}}_l \quad (9)$$

where $J(\hat{\mathbf{x}})$ is the Jacobian of the transformation from $\tilde{\boldsymbol{\varepsilon}}$ to (x_{l+1}, \dots, x_J) when $(x_1, \dots, x_l) = 0$.

Suppose the researcher has a data sample with $i = 1, \dots, N$ independent consumer purchase observations. The sample likelihood is

$$\mathcal{L}(\theta | \hat{\mathbf{x}}) = \prod_{i=1}^N f_x(\hat{\mathbf{x}}_i). \quad (10)$$

A maximum likelihood estimate of θ based on (10) is consistent and asymptotically efficient.

3.1.1 Estimation Challenges with the Neoclassical Model

van Soest, Kapteyn, and Kooreman (1993) have shown that the choice of functional form to approximate utility, $U(\mathbf{x})$, can influence consistency of the maximum likelihood estimator based on (10). In particular, the KKT conditions in (2) generate a unique vector $\mathbf{x}^*(\mathbf{p}, y; \theta, \boldsymbol{\varepsilon})$ at given (\mathbf{p}, y) for all possible θ and $\boldsymbol{\varepsilon}$ as long as $U(\mathbf{x})$ is monotonic and strictly quasi-concave. When these conditions fail to hold, the system of KKT conditions (2) may not generate a unique solution, $\mathbf{x}^*(\mathbf{p}, y; \theta, \boldsymbol{\varepsilon})$. This non-uniqueness leads to the well-known coherency problem with maximum

likelihood estimation (Heckman, 1978)¹¹, which can lead to inconsistent estimates. Note that the term coherency is used slightly differently in the more recent literature on empirical games with multiple equilibria. Tamer (2003) uses the term *coherency* in reference to the sufficient conditions for the existence of a solution $\mathbf{x}^*(\mathbf{p}, y; \theta, \varepsilon)$ to the model (in this case \mathbf{x}^* satisfies the KKT conditions). He uses the term *model completeness* in reference to the case where these sufficient conditions for the statistical model to have a well-defined likelihood. For our neoclassical model of demand, the econometric model would be termed “incomplete” if demand was a correspondence and, hence, there were multiple values of \mathbf{x}^* that satisfy the KKT conditions at a given $(\mathbf{p}, y; \theta, \varepsilon)$.

van Soest, Kapteyn, and Kooreman (1993) propose a set of parameter restrictions that are sufficient for coherency. For many specifications, these conditions will only ensure that the regularity of $U(x)$ holds over the set of prices and quantities observed in the data. While these conditions may suffice for estimation, failure of the global regularity condition could be problematic for policy simulations that use the demand estimates to predict outcomes outside the range of observed values in the sample. For many specifications, the parameter restrictions may not have an analytic form, and may require numerical tools to impose them. As we will see in the examples below, the literature has often relied on special functional forms, with properties like additivity and homotheticity, to ensure global regularity and to satisfy the coherency conditions. However, these specifications come at the cost of less flexible substitution patterns.

In addition to coherency concerns, maximum likelihood estimation based on equation (9) also involves several computational challenges. If the system of KKT conditions does not generate a closed-form expression for the conditional demand equations, it may be difficult to impose coherency conditions. In addition, the likelihood comprises a density component for the goods with non-zero consumption and a mass component for the corners at which some of the goods have an optimal demand of zero. The mass component in (9) requires evaluating an l -dimensional integral over a region defined implicitly by the solution to the KKT conditions (4). When consumers pur-

¹¹Coherency pertains to the case where there is a unique vector \mathbf{x}^* generated by the KKT conditions corresponding to each possible value of ε , and there is a unique value of ε that generates each possible vector \mathbf{x}^* generated by the KKT conditions.

chase l of the alternatives, there are $\binom{J}{l}$ potential shopping baskets, and each of the observed combinations would need to be solved. The likelihood also involves two change-of-variables from ε to $\tilde{\varepsilon}$ and from $\tilde{\varepsilon}$ to \hat{x} respectively, requiring the computation of a Jacobian matrix.

Estimation methods are beyond the scope of this discussion. However, a number of papers have proposed methods to accommodate several of the computational challenges above including simulated maximum likelihood (Kao, fei Lee, and Pitt, 2001), hierarchical Bayesian algorithms that use MCMC methods based on Gibbs sampling (Millimet and Tchernis, 2008), hybrid methods that combine Gibbs sampling with Metropolis-Hastings (Kim, Allenby, and Rossi, 2002), and GMM estimation (Thomassen, Seiler, Smith, and Schiraldi, 2017).

In the remainder of this section, we discuss several examples of functional forms for $U(x)$ that have been implemented in practice.

3.1.2 Example: Quadratic Utility

Due to its tractability, the quadratic utility,

$$U(\mathbf{x}; \theta, \varepsilon) = \sum_{j=1}^{J+1} (\beta_{j0} + \varepsilon_j) x_j + \frac{1}{2} \sum_{j=1}^{J+1} \sum_{k=1}^{J+1} \beta_{jk} x_j x_k \quad (11)$$

has been a popular functional form for empirical work (e.g., Wales and Woodland, 1983; Ransom, 1987; Lambrecht, Seim, and Skiera, 2007; Mehta, 2015; Yao, Mela, Chiang, and Chen, 2012; Thomassen, Seiler, Smith, and Schiraldi, 2017)¹². The random utility shocks in (11) are “random coefficients” capturing heterogeneity across consumers in the linear utility components over the various products. Assume WLOG that the consumer foregoes consumption on goods $x_j = 0$ ($j = 1, \dots, l$), and chooses a positive quantity for goods $x_j > 0$ ($j = l + 1, \dots, J + 1$). The corresponding KKT conditions are

¹²Thomassen, Seiler, Smith, and Schiraldi (2017) extend the quadratic utility model to allow for discrete store choice as well as the discrete/continuous expenditure allocation decisions across grocery product categories within a visited store.

$$\tilde{\varepsilon}_j + \beta_{j0} + \sum_{k=1}^{J+1} \beta_{jk} x_j^* - (\beta_{J+1,0} + \sum_{k=1}^{J+1} \beta_{J+1,k} x_{J+1}^*) p_j \leq 0 \quad , \quad j = 1, \dots, l \quad (12)$$

$$\tilde{\varepsilon}_j + \beta_{j0} + \sum_{k=1}^{J+1} \beta_{jk} x_j^* - (\beta_{J+1,0} + \sum_{k=1}^{J+1} \beta_{J+1,k} x_{J+1}^*) p_j = 0 \quad , \quad j = l+1, \dots, J$$

where $\tilde{\varepsilon}_j = \varepsilon_j - p_j \varepsilon_{J+1}$ and, by the symmetry condition, $\beta_{jk} = \beta_{kj}$. Since the quadratic utility function is homogeneous of degree zero in the parameters, we impose the normalization $\sum_{j=1}^{J+1} \beta_{j0} = 1$. We have also re-written the estimation problem in terms of differences, $\tilde{\varepsilon}$ to resolve the adding-up condition.

If $\tilde{\varepsilon} \sim N(0, \tilde{\Sigma})$ and the consumer purchases $\hat{\mathbf{x}} = (0, \dots, 0, \hat{x}_{l+1}, \dots, \hat{x}_J)$, the corresponding likelihood is¹³

$$f_{\mathbf{x}}(\hat{\mathbf{x}}) = \int_{-\infty}^{h_l(\hat{\mathbf{x}}; \theta)} \cdots \int_{-\infty}^{h_1(\hat{\mathbf{x}}; \theta)} f_{\varepsilon}(\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_l, h_{l+1}(\hat{\mathbf{x}}; \theta), \dots, h_J(\hat{\mathbf{x}}; \theta)) |J(\hat{\mathbf{x}})| d\tilde{\varepsilon}_1 \cdots d\tilde{\varepsilon}_l \quad (13)$$

where $h(\hat{\mathbf{x}}; \theta) = -\beta_{j0} - \sum_{k=1}^{J+1} \beta_{ki} x_j^* + (\beta_{J+1,0} + \sum_{k=1}^{J+1} \beta_{J+1,k} x_{J+1}^*) p_j$, $f_{\varepsilon}(\varepsilon)$ is the density corresponding to $N(0, \Sigma)$ and $J(\hat{x})$ is the Jacobian from $\tilde{\varepsilon}$ to (x_{l+1}, \dots, x_J) .

Ransom (1987) showed that the concavity of the quadratic utility function, (11), is sufficient for coherency of the maximum likelihood problem (13), even though monotonicity may not hold globally. Concavity is ensured if the matrix of cross-price effects, B where $B_{jk} = \beta_{jk}$, is symmetric and negative definite.

The advantages of the quadratic utility function include the flexibility of the substitution patterns between the goods, including a potential blend of complements and substitutes. However, the specification does not scale well in the number of products J . The number of parameters increases quadratically with J due to the cross-price effects. Moreover, the challenges in imposing global regularity could be problematic for policy simulations using the demand parameters.

¹³The density $f_{\varepsilon}(\tilde{\varepsilon})$ is induced by $f(\varepsilon)$ and the fact that $\tilde{\varepsilon}_j = \varepsilon_j - p_j \varepsilon_{J+1}$.

3.1.3 Example: Linear Expenditure System (LES)

One of the classic utility specifications in the demand estimation literature is the Stone-Geary model:

$$U(\mathbf{x}; \boldsymbol{\theta}) = \sum_{j=1}^{J+1} \theta_j \ln(x_j - \theta_{j1}), \quad \theta_j > 0. \quad (14)$$

Similar to the CES specification, the parameters θ_j measure the curvature of the sub-utility of each product and affect the rate of satiation. The translation parameters θ_{j1} allow for potential corner solutions.

The Stone-Geary preferences have been popular in the extant literature because the corresponding demand system can be solved analytically:

$$x_j^* = \theta_{j1} - \tilde{\theta}_j \sum_{k=1}^{J+1} \theta_{k1} \frac{p_k}{p_j} + \tilde{\theta}_j \frac{y}{p_j}, \quad j = 1, \dots, J+1 \quad (15)$$

where $x_j^* > \theta_{j1}$, $\forall j$, and where $\tilde{\theta}_j = \frac{\theta_j}{\sum_k \theta_k}$. The specification is often termed the ‘‘linear expenditure system’’ (LES) because the expenditure model is linear in prices

$$p_j x_j^* = \theta_{j1} p_j + \tilde{\theta}_j \left(y - \sum_k \theta_{k1} p_k \right). \quad (16)$$

Corner solutions with binding non-negativity constraints can arise when $\theta_{j1} \leq 0$ and, consequently, product j is ‘‘inessential’’ (Kao, fei Lee, and Pitt, 2001; Du and Kamakura, 2008). Assume WLOG that the consumer foregoes consumption on goods $x_j = 0$ ($j = 1, \dots, l$), and chooses a positive quantity for goods $x_j > 0$ ($j = l+1, \dots, J+1$). If we let $\theta_j = e^{\bar{\theta}_j + \varepsilon_j}$ where $\bar{\theta}_{J+1} = 0$ then the KKT conditions are:

$$\tilde{\varepsilon}_j + \bar{\theta}_j - \ln(-\theta_{j1}) + \ln(y - \sum_{k=1}^J x_k^* p_k - \theta_{J+1,1}) - \ln p_j \leq 0 \quad , j = 1, \dots, l \quad (17)$$

$$\tilde{\varepsilon}_j + \bar{\theta}_j - \ln(x_j^* - \theta_{j1}) + \ln(y - \sum_{k=1}^J x_k^* p_k - \theta_{J+1,1}) - \ln p_j = 0 \quad , j = l+1, \dots, J$$

where $\tilde{\varepsilon}_j = \varepsilon_j - \varepsilon_{J+1}$ and $\theta_{j1} \leq 0$ for $j = 1, \dots, l$.

If $\tilde{\varepsilon} \sim N(0, \Sigma)$ the consumer purchases $\hat{\mathbf{x}} = (0, \dots, 0, \hat{x}_{l+1}, \dots, \hat{x}_J)$, the corresponding likelihood is

$$f_x(\hat{\mathbf{x}}; \theta) = \int_{-\infty}^{h_l(\hat{\mathbf{x}}; \theta)} \cdots \int_{-\infty}^{h_1(\hat{\mathbf{x}}; \theta)} f_\varepsilon(\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_l, h_{l+1}(\hat{\mathbf{x}}; \theta), \dots, h_J(\hat{\mathbf{x}}; \theta)) |J(\hat{\mathbf{x}})| d\tilde{\varepsilon}_1 \cdots d\tilde{\varepsilon}_l \quad (18)$$

where $h_j(\hat{\mathbf{x}}; \theta) = -\bar{\theta}_j + \ln(-\theta_{j1}) - \ln(y - \sum_{k=1}^J x_k^* p_k - \theta_{J+1,1}) + \ln p_j$ and $f_\varepsilon(\varepsilon)$ is the density corresponding to $N(0, \Sigma)$ and $J(\hat{\mathbf{x}})$ is the Jacobian from $\tilde{\varepsilon}$ to (x_{l+1}, \dots, J) .

Some advantages of the LES specification include the fact that the utility function is globally concave, obviating the need for additional restrictions to ensure model coherency. In addition, the LES scales better than the quadratic utility as the number of parameters to be estimated grows linearly with the number of product, J . However, the specification does not allow for the same degree of flexibility in the substitution patterns between goods. The additive separability of the sub-utility functions associated with each good implies that the marginal utility of one good is independent of the level of consumption of all the other goods. Therefore, the goods are assumed to be strict Hicksian substitutes and any substitution between products arises through the budget constraint. The additive structure also rules out the possibility of inferior goods (see Deaton and Muellbauer, 1980b, page 139).

3.1.4 Example: Translated CES Utility

Another popular specification for empirical work is the translated CES utility function (Pollak and Wales, 1992; Kim, Allenby, and Rossi, 2002):

$$U(\mathbf{x}^*; \theta, \varepsilon) = \sum_{j=1}^J \psi_j (x_j + \gamma_j)^{\alpha_j} + \psi_{J+1} x_{J+1}^{\alpha_{J+1}} \quad (19)$$

where $\psi_j = \bar{\psi}_j \exp(\varepsilon_j) > 0$ is the stochastic perceived quality of a unit of product j , $\gamma_j \geq 0$ is a translation of the utility, $\alpha_j \in (0, 1]$ is a satiation parameter, and the collection of parameters to be estimated consists of $\theta = \{\alpha_j, \gamma_j, \bar{\psi}_j\}_{j=1}^{J+1}$. This specification nests several well-known models such as the translated Cobb-Douglas or “linear expenditure system” ($\alpha_j \rightarrow 0$) and the translated

Leontieff ($\alpha_j \rightarrow -\infty$). For any product j , setting $\gamma_j = 0$ would ensure a strictly interior quantity, $x_j^* > 0$. The CES specification has also been popular due to its analytic solution when quantities demanded are strictly interior. See for instance applications to nutrition preferences by Dubois, Griffith, and Nevo (2014) and Allcott, Diamond, Dube, Handbury, Rahkovsky, and Schnell (2018).

For the more general case with corner solutions, the logarithmic form of the KKT conditions associated with the translated CES utility model are

$$\tilde{\varepsilon}_j \leq h_j(x_j^*; \theta) \quad , \quad j = 1, \dots, l \quad (20)$$

$$\tilde{\varepsilon}_j = h_j(x_j^*; \theta) \quad , \quad j = l+1, \dots, J$$

where $h_j(x_j^*; \theta) = \ln\left(\bar{\psi}_{J+1}\alpha_{J+1}(x_j^*)^{\alpha_{J+1}-1}\right) - \ln\left(\bar{\psi}_j\alpha_j(x_j^* + \gamma_j)^{\alpha_j-1}\right) + \ln(p_j)$ and $\tilde{\varepsilon}_j = \varepsilon_j - \varepsilon_{J+1}$.

If $\tilde{\varepsilon} \sim N(0, \Sigma)$ and the consumer purchases $\hat{\mathbf{x}} = (0, \dots, 0, \hat{x}_{l+1}, \dots, \hat{x}_J)$, the corresponding likelihood is

$$f_x(\hat{\mathbf{x}}; \theta) = \int_{-\infty}^{h_l(\hat{\mathbf{x}}; \theta)} \cdots \int_{-\infty}^{h_1(\hat{\mathbf{x}}; \theta)} f_\varepsilon(\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_l, h_{l+1}(\hat{\mathbf{x}}; \theta), \dots, h_J(\hat{\mathbf{x}}; \theta)) |J(\hat{\mathbf{x}})| d\tilde{\varepsilon}_1 \cdots d\tilde{\varepsilon}_l. \quad (21)$$

where $f_\varepsilon(\varepsilon)$ is the density corresponding to $N(0, \Sigma)$ and $J(\hat{\mathbf{x}})$ is the Jacobian from $\tilde{\varepsilon}$ to (x_{l+1}, \dots, J) .

If instead we assume $\varepsilon \sim i.i.d. EV(0, \sigma)$, Bhat (2005) and Bhat (2008) derive the simpler, closed-form expression for the likelihood with analytic solutions to the integrals and the Jacobian $J(\hat{\mathbf{x}})$ in (21)

$$f_x(\hat{\mathbf{x}}; \theta) = \frac{1}{\sigma^{J-l}} \left[\prod_{i=l+1}^{J+1} f_i \right] \left[\sum_{i=l+1}^{J+1} \frac{p_i}{f_i} \right] \left[\frac{\prod_{i=l+1}^{J+1} e^{\frac{h_i(\hat{x}_i^*; \theta)}{\sigma}}}{\left(\sum_{j=1}^{J+1} e^{\frac{h_j(\hat{x}_j^*; \theta)}{\sigma}} \right)^{J-l+1}} \right] (J-l)! \quad (22)$$

where, changing the notation from above slightly, we define $h_j(x_j^*; \theta) = \bar{\psi}_j + (\alpha_j - 1) \ln(x_j^* + \gamma_j) -$

$\ln(p_j)$ and $f_i = \left(\frac{1-\alpha_i}{x_i^* + \gamma_i}\right)$.

A formulation of the utility function specifies an additive model of utility over stochastic consumption needs instead of over products (Hendel, 1999; Dubé, 2004)

$$U(\mathbf{x}^*; \theta, \psi) = \sum_{t=1}^T \left(\sum_{j=1}^J \psi_{jt} x_{jt} \right)^\alpha.$$

One interpretation is that the consumer shops in anticipation of T separate future consumption occasions (Walsh, 1995) where $T \sim \text{Poisson}(\lambda)$. The consumer draws the marginal utilities per unit of each product independently across the T consumption occasions, $\psi_{jt} \sim F(\bar{\psi}_j)$. The estimable parameters consist of $\theta = (\lambda, \bar{\psi}_1, \dots, \bar{\psi}_J, \alpha)$. For each of the $t = 1, \dots, T$ occasions, the consumer has perfect substitutes preferences over the products and chooses a single alternative. The purchase of variety on a given trip arises from the aggregation of the choices for each of the consumption occasions. Non-purchase is handled by imposing indivisibility on the quantities; although a translation parameter like the one in (19) above could also be used if divisibility was allowed.

Like the LES specification, the translated CES model is monotonic and quasi-concave, ensuring the consistency of the likelihood. The model also scales better than the quadratic utility as the number of parameters to be estimated grows linearly with the number of products, J . Scalability is improved even further by projecting the perceived quality parameters, ψ_j , onto a lower-dimensional space of observed product characteristics (Hendel, 1999; Dubé, 2004; Kim, Allenby, and Rossi, 2007). But, the translated CES specification derived above assumes the products are strict Hicksian substitutes, which limits the substitution patterns implied by the model.¹⁴ Moreover, with a large number of goods and small budget shares, the implied cross-price elasticities will be small in these models (Mehta, 2015).

¹⁴Kim, Allenby, and Rossi (2002) apply the model to the choices between flavor variants of yogurt where consumption complementarities are unlikely. However, this restriction would be more problematic for empirical studies of substitution between broader commodity groups.

3.1.5 Virtual Prices and the Dual Approach

Thus far, we have used a primal approach to derive the neoclassical model of demand with binding non-negativity constraints from a parametric model of utility. Most of the functional forms used to approximate utility in practice impose restrictions motivated by technical convenience. As we saw above, these restrictions can limit the flexibility of the demand model on factors such as substitution patterns between products and income effects. For instance, the additivity assumption resolves the global regularity concerns, but restricts the products to be strict substitutes. Accommodating more flexible substitution patterns becomes computationally difficult even for a simple specification like quadratic utility due to the coherency conditions.

The dual approach has been used to derive demand systems from less restrictive assumptions Deaton and Muellbauer (1980b). Lee and Pitt (1986) developed an approach to use duality to derive demand while accounting for with binding non-negativity constraints using virtual prices. The advantage of the dual approach is that a flexible functional form can be used to approximate indirect utility and cost functions can be used to determine the relevant restrictions to ensure that the derived demand system is consistent with microeconomic principles. The trade-off from using this dual approach is that the researcher loses the direct connection between the demand parameters and their deep structural interpretation as specific aspects of “preferences.” The specifications may be less suitable for marketing applications to problems such as product design, consumer quality choice and the valuation of product features, these specifications.

We begin with the consumer’s indirect utility function

$$V(\mathbf{p}, y; \boldsymbol{\theta}, \varepsilon) = \max_x \{U(\mathbf{x}; \boldsymbol{\theta}, \varepsilon) \mid \mathbf{p}'\mathbf{x} = y\} \quad (23)$$

where the underlying utility function $U(\mathbf{x}; \boldsymbol{\theta}, \varepsilon)$ is again assumed to be strictly quasi-concave, continuously differentiable and increasing. Roy’s Identity generates a system of notional demand

equations

$$\tilde{x}_j(\mathbf{p}, y; \boldsymbol{\theta}, \varepsilon) = -\frac{\frac{\partial V(\mathbf{p}, y; \boldsymbol{\theta}, \varepsilon)}{\partial p_j}}{\frac{\partial V(\mathbf{p}, y; \boldsymbol{\theta}, \varepsilon)}{\partial y}}, \forall j. \quad (24)$$

These demand equations are notional because they do not impose non-negativity and can therefore allow for negative values. In essence, $\tilde{\mathbf{x}}$ is a latent variable since it is negative for products that are not purchased. Note that Roy's identity requires that prices are fixed and independent of the quantities purchased by the consumer, an assumption that fails in settings where firms use non-linear pricing such as promotional quantity discounts¹⁵.

Lee and Pitt (1986) use virtual prices to handle products with zero quantity demanded (Neary and Roberts, 1980). Suppose the consumer's optimal consumption vector is $\mathbf{x}^* = (0, \dots, 0, x_{l+1}^*, \dots, x_{J+1}^*)$ where, as before, she does not purchase the first l goods. We can define virtual prices based on Roy's Identity in equation (24) that exactly set the notional demands to zero for the non-purchased goods

$$0 = \frac{\partial V(\boldsymbol{\pi}(\bar{\mathbf{p}}, y; \boldsymbol{\theta}, \varepsilon), \bar{\mathbf{p}}, y; \boldsymbol{\theta}, \varepsilon)}{\partial p_j}, \quad j = 1, \dots, l.$$

where $\boldsymbol{\pi}(\bar{\mathbf{p}}, y; \boldsymbol{\theta}, \varepsilon) = (\pi_1(\bar{\mathbf{p}}, y; \boldsymbol{\theta}, \varepsilon), \dots, \pi_l(\bar{\mathbf{p}}, y; \boldsymbol{\theta}, \varepsilon))$ is the l -vector of virtual prices and $\bar{\mathbf{p}} = (p_{l+1}, \dots, p_J)$. These virtual prices act like reservation prices for the non-purchased goods. We can derive the positive demands for goods $j = l + 1, \dots, J + 1$ by substituting the virtual prices into Roy's identity:

$$x_j^*(\bar{\mathbf{p}}, y; \boldsymbol{\theta}, \varepsilon) = -\frac{\frac{\partial V(\boldsymbol{\pi}(\bar{\mathbf{p}}, y; \boldsymbol{\theta}, \varepsilon), \bar{\mathbf{p}}, y; \boldsymbol{\theta}, \varepsilon)}{\partial p_j}}{\frac{\partial V(\boldsymbol{\pi}(\bar{\mathbf{p}}, y; \boldsymbol{\theta}, \varepsilon), \bar{\mathbf{p}}, y; \boldsymbol{\theta}, \varepsilon)}{\partial y}}, \quad j = l + 1, \dots, J + 1. \quad (25)$$

The regime switching conditions in which products $j = 1, \dots, l$ are not purchased consist of comparing virtual prices and observed prices:

$$\pi_j(\bar{\mathbf{p}}, y; \boldsymbol{\theta}, \varepsilon) \leq p_j, \quad j = 1, \dots, l. \quad (26)$$

Lee and Pitt (1986) demonstrate the parallel between the switching conditions based on virtual

¹⁵Howell, Lee, and Allenby (2016) show how a primal approach with a parametric utility specification can be used in the presence of non-linear pricing.

prices in (26) and the binding non-negativity constraints in the KKT conditions, (4).

The demand parameters θ can then be estimated by combining the conditional demand system, (25), and the regime-switching conditions, (4). If the consumer purchases $\hat{\mathbf{x}} = (0, \dots, 0, \hat{x}_{l+1}, \dots, \hat{x}_{J+1})$, the corresponding likelihood is

$$f_x(\hat{\mathbf{x}}; \theta) = \int_{\pi_1^{-1}(\mathbf{p}, y, p_l; \theta)}^{\infty} \cdots \int_{\pi_1^{-1}(\mathbf{p}, y, p_1; \theta)}^{\infty} f_{\varepsilon}(\varepsilon_1, \dots, \varepsilon_l, x_{l+1}^{*-1}(\hat{\mathbf{x}}, \bar{\mathbf{p}}, y; \theta), \dots, x_J^{*-1}(\hat{\mathbf{x}}, \bar{\mathbf{p}}, y; \theta)) |J(\hat{\mathbf{x}})| d\tilde{\varepsilon}_1 \cdots d\tilde{\varepsilon}_l. \quad (27)$$

where $f_{\varepsilon}(\varepsilon)$ is the density corresponding to $N(0, \Sigma)$ and $J(\hat{\mathbf{x}})$ is the Jacobian from $\tilde{\varepsilon}$ to (x_{l+1}, \dots, x_J) .

The inverse functions in (27) reflect the fact that $\pi_j^{-1}(\mathbf{p}, y, p_j; \theta) \leq \varepsilon_j$ for $j = 1, \dots, l$ and $x_j^{*-1}(\hat{\mathbf{x}}, \bar{\mathbf{p}}, y; \theta) = \varepsilon_j$ for $j = l + 1, \dots, J$.

As with the primal problem, the choice of functional form for the indirect utility, $V(\mathbf{p}, y; \theta, \varepsilon)$, can influence the coherency of the maximum likelihood estimator for θ in (27). van Soest, Kapteyn, and Kooreman (1993) show that the uniqueness of the demand function defined by Roy's Identity in (24) will hold if the indirect utility function $V(\mathbf{p}, y; \theta, \varepsilon)$ satisfies the following three regularity conditions:

1. $V(\mathbf{p}, y; \theta, \varepsilon)$ is homogeneous of degree zero
2. $V(\mathbf{p}, y; \theta, \varepsilon)$ is twice continuously differentiable in p and y
3. $V(\mathbf{p}, y; \theta, \varepsilon)$ is regular, meaning that the Slutsky matrix is negative semi-definite (“negativity”)

For many popular and convenient flexible functional forms, the Slutsky matrix may fail to satisfy negativity leading to the coherency problem. In many of these cases, such as AIDS, the virtual prices may need to be derived numerically, making it difficult to derive analytic restrictions that would ensure these regularity conditions hold. For these reasons, the homothetic translog specification discussed below has been extremely popular in practice.

3.1.6 Example: Indirect Translog Utility

One of the most popular implementations of the dual approach described above in section (3.1.5) uses the translog approximation of the indirect utility function (e.g., Lee and Pitt, 1986; Millimet and Tchernis, 2008; Mehta, 2015)

$$V(\mathbf{p}, y; \boldsymbol{\theta}, \boldsymbol{\varepsilon}) = \sum_{j=1}^{J+1} \theta_j \ln \left(\frac{p_j}{y} \right) + \frac{1}{2} \sum_{j=1}^{J+1} \sum_{k=1}^{J+1} \theta_{jk} \ln \left(\frac{p_j}{y} \right) \ln \left(\frac{p_k}{y} \right). \quad (28)$$

The econometric error is typically introduced by assuming $\theta_j = \bar{\theta}_j + \varepsilon_j$ where $\varepsilon_j \sim F(\boldsymbol{\varepsilon})$.

Roy's Identity gives us the notional expenditure share for product j CHE

$$s_j = \frac{-\theta_j - \sum_{k=1}^{J+1} \theta_{jk} \ln \frac{p_k}{y}}{1 - \sum_k \sum_l \theta_{kl} \ln \left(\frac{p_l}{y} \right)}. \quad (29)$$

van Soest and Kooreman (1990) derived slightly weaker sufficient conditions for coherency of the translog approach than van Soest, Kapteyn, and Kooreman (1993). Following van Soest and Kooreman (1990), we impose the following additional restrictions which are sufficient for the concavity of the underlying utility function and, hence, the uniqueness of the demand system (29) for a given realization of $\boldsymbol{\varepsilon}$:

$$\theta_{J+1} = 1 - \sum_j \theta_j$$

$$\sum_k \theta_{jk} = 0, \forall j \quad (30)$$

$$\theta_{jk} = \theta_{kj}, \forall j$$

We can re-write the expenditure share for product j

$$s_j = -\theta_j - \sum_{k=1}^J \theta_{jk} \ln p_k. \quad (31)$$

We can see from (31) that an implication of the restrictions in (30) is that they also impose homotheticity on preferences. For the translog specification, Mehta (2015) derived necessary and

sufficient conditions for global regularity that are even weaker than the conditions in van Soest and Kooreman (1990). These conditions allow for more flexible income effects (normal and inferior goods) and for more flexible substitution patterns (substitutes and complements), mainly by relaxing homotheticity.¹⁶

3.2 The Discrete/Continuous Product Choice Restriction in the Neoclassical Model

Perhaps due to their computational complexity, the application of the microeconomic models of variety discussed in section (3) has been limited. However, the empirical regularities documented in section (2) suggest that simpler models of discrete choice, with only a single product being chosen, could be used in many settings. Recall from section (2) that the average category has a single product choice chosen during 97% of trips. We now examine how our demand framework simplifies under discrete product choice. The discussion herein follows Hanemann (1984); Chiang and Lee (1992); Chiang (1991); Chintagunta (1993).

3.2.1 The Primal Problem

The model in this section closely follows the *simple re-packaging model with varieties* from Deaton and Muellbauer (1980b). In these models, the consumption utility for a given product is based on its effective quantity consumed, which scales the quantity by the product's quality. As before, we assume the commodity group of interest comprises $j = 1, \dots, J$ substitute products. Products are treated as perfect substitutes so that, at most, a single variant is chosen. We also assume there is an additional essential numeraire good indexed as product $J + 1$.

To capture discrete product choice within the commodity group, we assume the following

¹⁶In an empirical application to consumer purchases over several CPG product categories, Mehta (2015) finds the proposed model fits the data better than the homothetic translog specification. However, when $J = 2$, the globally regular translog will exhibit the restrictive strict Hicksian substitutes property.

bivariate utility over the commodity group and the essential numeraire:

$$U(\mathbf{x}^*; \theta, \boldsymbol{\psi}) = \tilde{U} \left(\sum_{j=1}^J \boldsymbol{\psi}_j x_j, \boldsymbol{\psi}_{J+1} x_{J+1} \right). \quad (32)$$

The parameter vector $\boldsymbol{\psi} = (\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_{J+1})$, $\boldsymbol{\psi}_j \geq 0$ measures the constant marginal utility of each of the products. In the literature, we often refer to $\boldsymbol{\psi}_j$ as the “perceived quality” of product j . Specifying the perceived qualities as random variables, $\boldsymbol{\psi} \sim F(\boldsymbol{\psi})$, introduces random utility as a potential source of heterogeneity across consumers in their perceptions of product quality. We also assume regularity conditions on $U(\mathbf{x}^*; \theta, \boldsymbol{\psi})$ to ensure that a positive quantity of x_{J+1} is always chosen.

To simplify the notation, let the total commodity vector be $z_1 = \sum_{j=1}^J \boldsymbol{\psi}_j x_j$ and let $z_2 = \boldsymbol{\psi}_{J+1} x_{J+1}$ so that we can re-write utility as $\tilde{U}(z_1, z_2)$. The KKT conditions are

$$\frac{\partial \tilde{U}(\boldsymbol{\psi}' x, \boldsymbol{\psi}_{J+1} x_{J+1})}{\partial z_1} \boldsymbol{\psi}_j - \frac{\partial \tilde{U}(\boldsymbol{\psi}' x, \boldsymbol{\psi}_{J+1} x_{J+1})}{\partial z_2} \boldsymbol{\psi}_{J+1} p_j \leq 0 \quad , \quad j = 1, \dots, J \quad (33)$$

where $\frac{\partial \tilde{U}(\boldsymbol{\psi}' x, \boldsymbol{\psi}_{J+1} x_{J+1})}{\partial z_1}$ is the marginal utility of total quality-weighted consumption within the commodity group. Because of the perfect substitutes specification, if a product within the commodity group is chosen, it will be product k if

$$\frac{p_k}{\boldsymbol{\psi}_k} = \min \left\{ \frac{p_j}{\boldsymbol{\psi}_j} \right\}_{j=1}^J$$

and, hence, k exhibits the lowest price-to-quality ratio. As with the general model in section (3), demand estimation will need to handle the regime switching, or demand selection conditions.

If $p_k > \frac{\frac{\partial \tilde{U}(z_1^*, z_2^*; \theta, \boldsymbol{\psi})}{\partial z_1} \boldsymbol{\psi}_k}{\frac{\partial \tilde{U}(z_1^*, z_2^*; \theta, \boldsymbol{\psi})}{\partial z_2} \boldsymbol{\psi}_{J+1}}$, then the consumer spends her entire budget on the numeraire: $x_{J+1}^* =$

y. Otherwise, the consumer allocates her budget between x_k^* and x_{J+1}^* to equate $\frac{\frac{\partial \tilde{U}(z_1^*, z_2^*; \theta, \boldsymbol{\psi})}{\partial z_1} \boldsymbol{\psi}_k}{p_k} = \frac{\frac{\partial \tilde{U}(z_1^*, z_2^*; \theta, \boldsymbol{\psi})}{\partial z_2} \boldsymbol{\psi}_{J+1}}{p_k}$.

We define $h_j(\mathbf{x}^*; \theta, \psi) = \frac{p_j \psi_{J+1} \frac{\partial \bar{u}(z_1^*, z_2^*; \theta, \psi)}{\partial z_2}}{\frac{\partial \bar{u}(z_1^*, z_2^*; \theta, \psi)}{\partial z_1}}$. When none of the products are chosen, we can write the likelihood of $\hat{\mathbf{x}} = (0, \dots, 0)$ as

$$f_x(\hat{\mathbf{x}}; \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_J(\hat{\mathbf{x}}; \theta, \psi_{J+1}) \cdots \int_{-\infty}^{\infty} h_1(\hat{\mathbf{x}}; \theta, \psi_{J+1}) f_\psi(\psi) d\psi_1 \cdots d\psi_{J+1}. \quad (34)$$

When product 1 (WLOG) is chosen, we can write the likelihood of $\hat{\mathbf{x}} = (\hat{x}_1, 0, \dots, 0)$ as

$$f_x(\hat{\mathbf{x}}; \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_J(\hat{\mathbf{x}}; \theta, \psi_{J+1}) \cdots \int_{-\infty}^{\infty} h_2(\hat{\mathbf{x}}; \theta, \psi_{J+1}) f_\psi(h_1(\hat{\mathbf{x}}; \theta, \psi), \psi_2, \dots, \psi_{J+1}) |J(\hat{\mathbf{x}})| d\psi_2 \cdots d\psi_{J+1} \quad (35)$$

where $J(\hat{\mathbf{x}})$ is the Jacobian from ψ_1 to \hat{x}_1 . The likelihood now comprises a density component for the chosen alternative $j = 1$, and a mass function for the remaining goods.

3.2.2 Example: Translated CES Utility

Recall the translated CES utility function presented in section (3.1.4)(Bhat, 2005; Kim, Allenby, and Rossi, 2002):

$$U(\mathbf{x}^*; \theta, \varepsilon) = \sum_{j=1}^J \psi_j (x_j + \gamma_j)^{\alpha_j} + \psi_{J+1} x_{J+1}^{\alpha_{J+1}}.$$

We can impose discrete product choice with the restrictions $\alpha_j = 1$, $\gamma_j = 0$ for $j = 1, \dots, J$, which gives us perfect substitutes utility over the brands

$$U(\mathbf{x}^*; \theta, \varepsilon) = \sum_{j=1}^J \psi_j x_j + \psi_{J+1} x_{J+1}^\alpha.$$

Let $\psi_j = \exp(\bar{\psi}_j \varepsilon_j)$, $j = 1, \dots, J$ and $\psi_{J+1} = \exp(\varepsilon_{J+1})$, where $\varepsilon_j \sim i.i.d. EV(0, \sigma)$ (Deaton and Muellbauer, 1980a; Bhat, 2008). When none of the products are chosen, we can write the

likelihood of $\hat{\mathbf{x}} = (0, \dots, 0)$ as

$$f_x(\hat{\mathbf{x}}) = \frac{1}{1 + \sum_{j=1}^J \exp\left(\frac{\bar{\psi}_j - \ln(p_j) - \ln(\alpha y^{(\alpha-1)})}{\sigma}\right)}$$

which is the multinomial logit model. If WLOG alternative 1 is chosen in the commodity group, the likelihood of $\hat{\mathbf{x}} = (\hat{x}_1, \dots, 0)$ is

$$\begin{aligned} f(\hat{\mathbf{x}}; \theta) &= \int_{-\infty}^{\infty} \frac{1}{\sigma p_1} \exp\left(-\frac{h_1(\psi_{J+1}; \hat{x}_1, \mathbf{p}, \theta)}{\sigma}\right) \exp\left[-\frac{1}{p_1} \exp\left(-\frac{h_1(\psi_{J+1}; \hat{x}_1, \mathbf{p}, \theta)}{\sigma}\right)\right] \left(\frac{\alpha-1}{y-\hat{x}_1}\right) f_\varepsilon(\varepsilon_{J+1}) d\varepsilon_{J+1} \\ &= \left(\frac{\alpha-1}{y-\hat{x}_1}\right) \frac{1}{\sigma} \int_{-\infty}^{\infty} \exp\left(\frac{-\varepsilon_{J+1}}{\sigma}\right) \sum_{j=1}^J \exp\left(\frac{V_j}{\sigma}\right) \prod_j \exp\left[-\exp\left(\frac{-\varepsilon_{J+1}}{\sigma}\right) \exp\left(\frac{V_j}{\sigma}\right)\right] f_\varepsilon(\varepsilon_{J+1}) d\varepsilon_{J+1} \end{aligned}$$

where

$$V_j \equiv \bar{\psi}_j - \ln(p_j)$$

and

$$h_j(\psi_{J+1}; \hat{x}_1, \mathbf{p}, \theta) = \ln\left(\psi_{J+1} \alpha (y - \hat{x}_1)^{\alpha-1}\right) + \ln(p_j) - \bar{\psi}_j$$

and

$$P_k \equiv P(\varepsilon_k + \bar{\psi}_k - \ln(p_k) \geq \varepsilon_j + \bar{\psi}_j - \ln(p_j), j = 1, \dots, J) = \frac{\exp\left(\frac{\bar{\psi}_k - \ln(p_k)}{\sigma}\right)}{\sum_{j=1}^J \exp\left(\frac{\bar{\psi}_j - \ln(p_j)}{\sigma}\right)}.$$

3.2.3 Example: the Dual Problem with Indirect Translog Utility

Muellbauer (1974) has shown that maximizing the simple re-packaging model utility function in (32) generates a bivariate indirect utility function of the form $V\left(\frac{p_k}{\psi_{ky}}, \frac{1}{\psi_{J+1y}}\right)$ when product k is the preferred product in the commodity group of interest. Following the template in Hanemann (1984), several empirical studies of discrete/continuous brand choice have been derived based

on the indirect utility function and the dual virtual prices. For instance, Chiang (1991) and Arora, Allenby, and Ginter (1998) use a second-order flexible translog approximation of the indirect utility function¹⁷:

$$V(\mathbf{p}, y; \theta, \varepsilon) = \theta_1 \ln \frac{p_k}{\psi_k y} + \theta_2 \ln \frac{1}{\psi_{J+1} y} + \frac{1}{2} \theta_{11} \left[\ln \left(\frac{p_k}{\psi_k y} \right) \right]^2 + \frac{1}{2} \theta_{12} \ln \frac{p_k}{\psi_k y} \ln \frac{1}{\psi_{J+1} y} + \frac{1}{2} \theta_{22} \left[\ln \frac{1}{\psi_{J+1} y} \right]^2 \quad (36)$$

where $\frac{p_k}{\psi_k} = \min_j \left\{ \frac{p_j}{\psi_j} \right\}$, $\psi_j = \exp(\bar{\psi}_j + \varepsilon_j)$ for $j = 1, \dots, J$, $\psi_{J+1} = \exp(\varepsilon_{J+1})$ and $\varepsilon_j \sim i.i.d. EV(0, \sigma)$.

To facilitate the exposition, we impose the following restrictions to ensure coherency. But, the restrictions lead to the homothetic translog specification which eliminates potentially interesting income effects in substitution between products (see the concerns discussed earlier in section 3.1.5):

$$\begin{aligned} \theta_1 + \theta_2 &= -1 \\ \theta_{11} + \theta_{12} &= 0 \\ \theta_{12} + \theta_{22} &= 0. \end{aligned} \quad (37)$$

Roy's Identity gives us the notional expenditure share for product k

$$s_k = -\theta_1 - \theta_{11} \ln \frac{p_k}{\psi_k} + \theta_{11} \ln \psi_{J+1}. \quad (38)$$

From (38), we see that $\varepsilon_1 = \frac{\hat{s}_1 + \theta_1 + \theta_{11} \ln(p_1) - \theta_{11} \psi_1 + \theta_{11} \varepsilon_{J+1}}{\theta_{11}}$. We can now compute the quality-adjusted virtual price (or reservation price) for purchase by setting (38) to zero: $R(\varepsilon_{J+1}; \hat{\mathbf{s}}) = \exp\left(-\frac{\theta_1 + \theta_{11} \varepsilon_{J+1}}{\theta_{11}}\right)$.

If none of the products are chosen, $\frac{p_k}{\psi_k} > R(\varepsilon_{J+1}; \hat{\mathbf{s}})$ and we observe $\hat{\mathbf{s}} = (0, \dots, 0)$ with likelihood

$$f(\hat{\mathbf{s}}; \theta) = \frac{\exp\left(\frac{\theta_1}{\sigma \theta_{11}}\right)}{\exp\left(\frac{\theta_1}{\sigma \theta_{11}}\right) + \sum_{j=1}^J \exp\left(\frac{\bar{\psi}_j - \ln(p_j)}{\sigma}\right)}. \quad (39)$$

¹⁷See Hanemann (1984) for other specification including LES and PIGLOG preferences. See also Chintagunta (1993) for the linear expenditure system or "Stone-Geary" specification.

If, WLOG, product 1 is chosen, the likelihood of $\hat{\mathbf{s}} = (\hat{s}_1, 0, \dots, 0)$ is

$$f(\hat{\mathbf{s}}) = \int_{-\infty}^{\infty} \frac{1}{\sigma \theta_{11} P_1} e^{-\frac{\hat{s}_1 + \theta_1 + \theta_{11} \ln(p_1) - \theta_{11} \psi_1 + \theta_{11} \varepsilon_{J+1}}{\theta_{11} \sigma}} e^{-\frac{1}{P_1}} e^{-\frac{\hat{s}_1 + \theta_1 + \theta_{11} \ln(p_1) - \theta_{11} \psi_1 + \theta_{11} \varepsilon_{J+1}}{\theta_{11} \sigma}} f_{\varepsilon}(\varepsilon_{J+1}) d\varepsilon_{J+1} \quad (40)$$

where

$$P_k \equiv P(\varepsilon_k + \bar{\psi}_k - \ln(p_k) \geq \varepsilon_j + \bar{\psi}_j - \ln(p_j), j = 1, \dots, J) = \frac{\exp\left(\frac{\bar{\psi}_k - \ln(p_k)}{\sigma}\right)}{\sum_{j=1}^J \exp\left(\frac{\bar{\psi}_j - \ln(p_j)}{\sigma}\right)}$$

As discussed in Mehta, Chen, and Narasimhan (2010), the distributional assumption $\varepsilon_j \sim$ i.i.d. $EV(0, \sigma)$ imposes a strong restriction on the price elasticity of the quantity purchased (conditional on purchase and brand choice), setting it very close to -1 . This property can be relaxed by using a more flexible distribution, such as multivariate normal errors. Alternatively, allowing for unobserved heterogeneity in the parameters of the conditional budget share expression (38) would alleviate this restriction at the population level.

3.2.4 Promotion Response: empirical findings using the discrete/continuous demand model

An empirical literature has used the discrete/continuous specification of demand to decompose the total price elasticity of demand into three components: (1) purchase incidence (2) brand choice and (3) quantity choice. This literature seeks to understand the underlying consumer choice mechanism that drives the observation of a large increase in quantities sold in response to a temporary price cut. In particular, the research assesses the extent to which a pure discrete brand choice analysis, focusing only on component (2) (see section 3.3 below), might miss part of the price elasticity and misinform the researcher or the retailer. Early work typically found that brand-switching elasticities accounted for most of the total price elasticity of demand in CPG product categories (e.g., Chiang, 1991; Chintagunta, 1993), though the unconditional brand choice elasticities were found to be larger than choice elasticities that condition on purchase. More recently, Bell, Chiang, and

Padmanabhan (1999)’s empirical generalizations indicate that the relative role of the brand switching elasticity varies across product categories. On average, they find that the quantity decision accounts for 25% of the total price elasticity of demand, suggesting that purchase acceleration effects may be larger than previously assumed. These results are based on static models in which any purchase acceleration would be associated with an increase in consumption. In section 5.1, we extend this discussion to models that allow for forward-looking consumers to stock-pile storable consumer goods in anticipation of higher future prices.

3.3 Indivisibility and the Pure Discrete Choice Restriction in the Neoclassical Model

The pure discrete choice behavior documented in the empirical stylized facts in section 2 suggests a useful restriction for our demand models. In many product categories, the consumer purchases at most one unit of a single product on a given trip. Discrete choice behavior also broadly applies to other non-CPG product domains such as automobiles, computers and electronic devices. The combination of pure discrete choice and indivisibility simplifies the discrete product choice model in section 3.2 by eliminating the intensive margin of quantity choice, reducing the model to one of pure product choice. Not surprisingly, pure discrete choice models have become extremely popular for modeling demand both in the context of micro data on consumer-level choices and with more macro data on aggregate market shares. We now discuss the relationship between the classic pure discrete choice models of demand estimated in practice (e.g. multinomial logit and probit) and contrast them to the derivation of pure discrete choice from the neoclassical models derived above.

3.3.1 A Neoclassical Derivation of the Pure Discrete Choice Model of Demand

Recall from section 3 where we defined the neoclassical economic model of consumer choice based on the following utility maximization problem:

$$V(\mathbf{p}, y; \theta, \varepsilon) \equiv \max_x \{U(\mathbf{x}; \theta, \varepsilon) : \mathbf{x}'\mathbf{p} \leq y, x \geq 0\} \quad (41)$$

where we assume $U(\bullet; \theta, \varepsilon)$ is a continuously-differentiable, quasi-concave and increasing function. In that problem, we assumed non-negativity and perfect divisibility, $x_j \geq 0$, for each of the $J = 1, \dots, J$ products and the $J + 1$ essential numeraire. We now consider the case of indivisibility on the $j = 1, \dots, J$ products by adding the restriction $x_j \in \{0, 1\}$ for $j = 1, \dots, J$. We also assume strong separability (i.e. additivity) of x_{J+1} and perfect substitutes preferences over the $j = 1, \dots, J$ products such that:

$$U\left(\sum_{j=1}^K \psi_j x_j, \psi_{J+1} x_{J+1}; \theta\right) = \sum_{j=1}^J \psi_j x_j + \tilde{u}(x_{J+1}; \Psi_{J+1})$$

where $\psi_j = \bar{\psi}_j + \varepsilon_j$ and $\tilde{u}(x_{j+1}; \bar{\psi}_{J+1}) = u(x_{j+1}; \bar{\psi}_{J+1}) + \varepsilon_{J+1}$. The KKT conditions will no longer hold under indivisible quantities. The consumer's choice problem consists of making a discrete choice among the following $J + 1$ choice-specific indirect utilities:

$$v_j = \bar{\psi}_j + u(y - p_j; \bar{\psi}_{J+1}) + \varepsilon_j + \varepsilon_{J+1} = \bar{v}_j + \varepsilon_j + \varepsilon_{J+1}, \quad x_j = 1 \quad (42)$$

$$v_{J+1} = u(y; \bar{\psi}_{J+1}) + \varepsilon_{J+1} = \bar{v}_{J+1} + \varepsilon_{J+1}, \quad x_{J+1} = y.$$

The probability that consumer chooses alternative $1 \in \{1, \dots, J\}$ is (WLOG)

$$\begin{aligned} Pr(x_1 = 1) &= Pr(v_1 \geq v_k, k \neq 1) \\ &= Pr(\varepsilon_k \leq \bar{v}_1 - \bar{v}_k + \varepsilon_1, \forall k \neq 1, \varepsilon_1 \geq \bar{v}_{J+1} - \bar{v}_1) \quad (43) \\ &= \int_{\bar{v}_{J+1} - \bar{v}_1}^{\infty} \int_{-\infty}^{\bar{v}_1 - \bar{v}_2 + x} \cdots \int_{-\infty}^{\bar{v}_1 - \bar{v}_J + x} f(x, \varepsilon_2, \dots, \varepsilon_J) d\varepsilon_J \cdots d\varepsilon_2 dx \end{aligned}$$

where $f(\varepsilon_1, \dots, \varepsilon_J)$ is the density of $(\varepsilon_1, \dots, \varepsilon_J)'$ and the probability of allocating the entire budget to the essential numeraire is simply $Pr(x_{J+1} = y) = 1 - \sum_{j=1}^J Pr(x_j = 1)$.

If we assume $(\varepsilon_1, \dots, \varepsilon_J)' \sim i.i.d. EV(0, 1)$, the choice probabilities in (43) become

$$Pr(x_1 = 1) = \frac{\exp(\bar{\psi}_1 + \bar{u}(y - p_1; \bar{\psi}_{J+1}))}{\sum_{k=1}^J \exp(\bar{\psi}_k + \bar{u}(y - p_k; \bar{\psi}_{J+1}))} \left[1 - e^{-e^{(-\bar{u}(y; \bar{\psi}_{J+1})) \sum_{k=1}^J e^{(\bar{\psi}_k + \bar{u}(y - p_k; \bar{\psi}_{J+1}))}} \right] \quad (44)$$

$$Prob(x_{J+1} = y) = \left[1 - e^{-e^{(-\bar{u}(y; \bar{\psi}_{J+1})) \sum_{k=1}^J e^{(\bar{\psi}_k + \bar{u}(y - p_k; \bar{\psi}_{J+1}))}} \right].$$

Suppose the researcher has a data sample with $i = 1, \dots, N$ independent consumer purchase observations. A maximum likelihood estimator of the model parameters can be constructed as follows:

$$\mathcal{L}(\theta | \mathbf{y}) = \prod_{i=1}^N \prod_{j=1}^J Pr(x_{J+1} = y)^{y_{ij+1}} Pr(x_j = 1)^{y_{ij}} \quad (45)$$

where y_{ij} indicates whether observation i resulted in choice alternative j , and $\theta = (\bar{\psi}_1, \dots, \bar{\psi}_{J+1})'$.

While the probabilities (44) generate a tractable maximum likelihood estimator based on (45), the functional forms do not correspond to the familiar multinomial logit specification used throughout the literature on discrete choice demand McFadden (e.g., 1981)¹⁸. To understand why the neoclassical models from earlier sections do not nest the usual discrete choice models, note that the random utilities ε_{J+1} “difference out” in (43) and the the model is equivalent to a deterministic utility for the decision to allocate the entire budget to the numeraire. This result arises because of the adding-up condition associated with the budget constraint, which we resolved by assuming $x_{J+1}^* > 0$, just as we did in section 3 above.

Lee and Allenby (e.g., 2014) extend the pure discrete choice model to allow for multiple discrete choice and indivisible quantities for each product. As before, assume there are $j = 1, \dots, J$ products and a $J + 1$ essential numeraire. To address the indivisibility of the $j = 1, \dots, J$ products, assume $x_j \in \{0, 1, \dots\}$ for $j = 1, \dots, J$. If utility is concave, increasing and additive¹⁹, $U(x) = \sum_{j=1}^J u_j(x_j) + \alpha_{J+1}(x_{J+1})$, the consumer’s decision problem consists of selecting an optimal quan-

¹⁸Besanko, Perry, and Spady (1990) study the monopolistic equilibrium pricing and variety of brands supplied in a market with discrete choice demand of the form 44.

¹⁹The tractability of this problem also requires assuming linearity in the essential numeraire $u(x_{J+1}) = \alpha(x_{J+1})$ so that the derivation of the likelihood can be computed separately for each product alternative.

tity for each of the products and the essential numeraire, subject to her budget constraint. Let $F = \{(x_1, \dots, x_J) \mid y - \sum_j x_j p_j \geq 0, x_j \in \{0, 1, \dots\}\}$ be the set of feasible quantities that satisfy the consumer's budget constraint, where $x_{J+1} = y - \sum_j x_j p_j$. The consumer picks an optimal quantity vector $\mathbf{x}^* \in F$ such that $U(\mathbf{x}^*) \geq U(x) \forall x \in F$.

To derive a tractable demand solution, Lee and Allenby (2014) assume that utility has the following form²⁰:

$$u_j(x) = \frac{\alpha_j \exp(\varepsilon_j)}{\gamma_j} \log(\gamma_j x + 1).$$

The additive separability assumption is critical since it allows the optimal quantity of each brand to be determined separately. In particular, for each $j = 1, \dots, J$ the optimality of x_j^* is ensured if $U(x_1^*, \dots, x_j^*, \dots, x_j^*) \geq \max\{U(x_1^*, \dots, x_j^* + \Delta, \dots, x_j^*) \mid \mathbf{x}^* \in F, \Delta \in \{-1, 1\}\}$. The limits of integration of the utility shocks, ε , can therefore be derived in closed form:

$$f_x(\hat{\mathbf{x}}; \theta) = \prod_{j=1}^J \int_{lb_j}^{ub_j} f_\varepsilon(\varepsilon_j) d\varepsilon_j$$

where $lb_j = \log\left(\frac{\alpha_{j+1} p_j \gamma_j}{\alpha_j}\right) - \log\left(\log\left(\frac{\gamma_j \hat{x}_{j+1}}{\gamma_j (\hat{x}_j - 1) + 1}\right)\right)$ and $ub_j = \log\left(\frac{\alpha_{j+1} p_j \gamma_j}{\alpha_j}\right) - \log\left(\log\left(\frac{\gamma_j (\hat{x}_j + 1) + 1}{\gamma_j \hat{x}_j + 1}\right)\right)$.

3.3.2 The Classic Pure Discrete Choice Model of Demand

Suppose as before that the consumer makes a discrete choice between each of the products in the commodity group. We again assume a bivariate utility over an essential numeraire and a commodity group, with perfect substitutes over the $j = 1, \dots, J$ products in the commodity group: $U\left(\sum_{j=1}^J \psi_j x_j, x_{J+1}\right)$. If we impose indivisibility on the product quantities such that $x_j \in \{0, 1\}$,

²⁰This specification is a special case of the translated CES model described earlier when the satiation parameter asymptotes to 0 (e.g., Bhat, 2008).

the choice problem once again becomes a discrete choice among the $j = 1, \dots, J + 1$ alternatives

$$v_j = U(\psi_j, y - p_j) + \varepsilon_j, \quad j = 1, \dots, J \quad (46)$$

$$v_{J+1} = U(0, y) + \varepsilon_{J+1}.$$

In this case, the random utility ε_{J+1} does not “difference out” and hence we will end up with a different system of choice probabilities. If we again assume that $(\varepsilon_1, \dots, \varepsilon_{J+1}) \sim i.i.d. EV(0, 1)$, the corresponding choice probabilities have the familiar multinomial logit (MNL) form:

$$\begin{aligned} Pr(j) &= Prob(v_j \geq v_k, \text{ for } k \neq j) \\ &= \frac{\exp(U(\psi_j, y - p_j))}{\exp(U(0, y)) + \sum_{k=1}^J \exp(U(\psi_k, y - p_k))}. \end{aligned} \quad (47)$$

Similarly, assuming $(\varepsilon_1, \dots, \varepsilon_{J+1}) \sim N(0, \Sigma)$ would give rise to the standard multinomial probit.

We now examine why we did not obtain the same system of choice probabilities as in the previous section. Unlike the derivation in the previous section, the random utilities in (46) were not specified as primitive assumptions on the underlying utility function $U\left(\sum_{j=1}^J \psi_j x_j, x_{J+1}\right)$. Instead, they were added on to the choice-specific values. An advantage of this approach is that it allows the researcher to be more agnostic about the exact interpretation of the errors. In the econometrics literature, ε are interpreted as unobserved product characteristics, unobserved utility or tastes, measurement error or specification error. However, the probabilistic choice model has also been derived by mathematical psychologists (e.g., Luce, 1977) who interpret the shocks as psychological states, leading to potentially non-rational forms of behavior. Whereas econometricians interpret the probabilistic choice rules in (47) as the outcome of utility maximization with random utility components, mathematical psychologists interpret (47) as stochastic choice behavior (see the discussion in Anderson, de Palma, and Thisse, 1992, chapters 2.4 and 2.5). A more recent literature has derived the multinomial logit from a theory of “rational inattention.” Under rational inattention, the stochastic component of the model captures a consumer’s product uncertainty and

the costs of endogenously reducing uncertainty through search (e.g., Matejka and McKay, 2015; Joo, 2018).

One approach to rationalize the system (47) is to define the $J + 1$ alternative as an additional non-market good with price $p_0 = 0$, usually defined as “home production” (e.g., Anderson and de Palma, 1992). We assume the consumer always chooses at least one of the $J + 1$ alternative. In addition, we introduce a divisible, essential numeraire good, z , with price $p_z = 1$, so that the consumer has bivariate utility over the total consumption of the goods and over the essential numeraire: $U\left(\sum_{j=0}^{J+1} \psi_j x_j, z\right)$. The choice-specific values correspond exactly to (46) and the shock ε_{J+1} is now interpreted as the random utility from home production. This model differs from the neoclassical models discussed in sections 3 and 3.2 because we have now included an additional non-market good representing household production.

For example, suppose a consumer has utility:

$$U\left(\sum_{j=0}^{J+1} \psi_j x_j, z\right) = \left(\sum_{j=0}^{J+1} \psi_j x_j\right) \exp(\alpha z)$$

where goods $j = 1, \dots, J + 1$ are indivisible, perfect substitutes each with perceived qualities $\psi_j = \exp(\bar{\psi}_j + \varepsilon_j)$, where we normalize $\bar{\psi}_{J+1} = 1$, and where α is preference for the numeraire good.. In this case, the choice-specific indirect utilities would be (in logs)

$$v_j = \psi_j + \alpha(y - p_j) + \varepsilon_j, \quad j = 1, \dots, J$$

$$v_{J+1} = \alpha y + \varepsilon_{J+1}.$$

The MNL was first applied to marketing panel data for individual consumers by Guadagni and Little (1983). The linearity of the conditional indirect utility function explicitly rules out income effects in the substitution patterns between the inside goods. We discuss tractable specifications that allow for income effects in section 4.1 below. If income is observed, then income effects in the substitution between the the commodity group and the essential numeraire can be incorporated by

allowing for non-linearity in the utility of the numeraire. For instance, if $\tilde{U}(x_{J+1}) = \psi_{J+1} \ln(x_{J+1})$ then we get choice probabilities²¹

$$Pr(k; \theta) = \frac{\exp(\psi_k + \psi_{J+1} \ln(y - p_k))}{\exp(\psi_{J+1} \ln(y)) + \sum_{j=1}^J \exp(\psi_j + \psi_{J+1} \ln(y - p_j))}. \quad (48)$$

This specification also imposes an affordability condition by excluding any alternative for which $p_j > y$.

The appeal of the MNL's closed-form specification comes at a cost for demand analysis. If $\tilde{U}(x_{J+1}) = \psi_{J+1} x_{J+1}$ as is often assumed in the literature, the model exhibits the well-known Independence of Irrelevant Alternatives property. The IIA property can impose unrealistic substitution patterns in demand analysis. At the individual consumer level, the cross-price elasticity of demand is constant:

$$\frac{\partial Pr(j)}{\partial p_k} \frac{p_k}{Pr(j)} = \psi_{J+1} Pr(k) p_k$$

so that substitution patterns between products will be driven by their prices and purchase frequencies, regardless of attributes. Moreover, a given product competes uniformly on price with all other products. One solution is to use a non-IIA specification. For instance, error components variants of the extreme value distribution, like nested logit and the generalized extreme value distribution, can relax the IIA property within pre-determined groups of products (e.g., McFadden, 1981; Cardell, 1997)²². If we instead assume that $\varepsilon \sim N(0, \Sigma)$ with appropriately scaled covariance matrix Σ , we obtain the multinomial probit (e.g., McCulloch and Rossi, 1994; Goolsbee and Petrin, 2004). Dotson, Howell, Brazell, Otter, Lenk, MacEachern, and Allenby (2018) parameterize the covariance matrix, Σ , using product characteristics to allow for a scalable model with correlated utility errors and, hence, stronger substitution between similar products. When consumer panel data are available, another solution is to use a random coefficients specification that allows for more flexible

²¹We can derive this specification from the assumption of Cobb-Douglas utility: $U(x_1, \dots, x_{J+1}; \theta) = \exp\left(\sum_{j=1}^{J+1} \psi_j x_j\right) x_{J+1}^{\psi_{J+1}}$.

²²Misra (2005) shows that the disutility minimization formulation of the multinomial logit (or "reverse logit") leads to a different functional form of the choice probabilities that does not exhibit the IIA property.

aggregate substitution patterns (see Chapter 2 of this volume).

In their seminal application of the multinomial logit to consumer-level scanner data, Guadagni and Little (1983) estimated demand for the ground coffee category using 78 weeks of transaction data for 2,000 households shopping in 4 Kansas City Supermarkets. Interestingly, they found that brand and pack size were the most predictive attributes for consumer choices. They also included the promotional variables “feature ad” and “in-aisle display” as additive utility shifters. These variables have routinely been found to be predictive of consumer choices. However, the structural interpretation of a marginal utility from a feature ad or a display is ambiguous. While it is possible that consumers obtain direct consumption value from a newspaper ad or a display, it seems more likely that these effects are the reduced-form of some other process such as information search. Exploring the structural foundations of the “promotion effects” remains a fruitful area for future research.

4 Some Extensions to the Typical Neoclassical Specifications

4.1 Income Effects

Most of the empirical specifications discussed earlier imposed regularity conditions that, as a byproduct, impose strong restrictions on the income effects on demand. Since the seminal work by Engel (1857), the income elasticity of demand has been used to classify goods based on consumption behavior. Goods with a positive income elasticity are classified as Normal goods, for which consumers increase their consumption as income increases. Goods with a negative income elasticity are classified as Inferior goods, for which consumers decrease their consumption as income increases. Engel’s law is based on the empirical observation that households tend to allocate a higher proportion of their income to food as they become poorer (e.g., Engel, 1857). Accordingly, we define *necessity* goods and *luxury* goods based on whether the income elasticity of demand is less than or greater than one. Homothetic preferences restrict all products to be strict Normal goods with an income elasticity of one, thereby limiting the policy implications one can study with the

model. Quasi-linear preferences over the composite “outside” good restrict the income elasticity to zero, eliminating income effects entirely.

When the empirical focus is on a specific product category for a low-priced item like a CPG product, it may be convenient to assume that income effects are likely to be small and inconsequential²³. This assumption is particularly convenient when a household’s income or shopping budget is not observed. However, overly restrictive income effects can limit a model’s predicted substitution patterns, leading to potentially adverse policy implications (see McFadden’s forward to Anderson, de Palma, and Thisse (1992) for a discussion). Even when a household’s income is static, large changes in relative prices could nevertheless create purchasing power effects. Consider the bivariate utility function specification with perfect substitutes in the focal commodity group from section 3.2: $\tilde{U} \left(\sum_{j=1}^J \psi_j x_j, x_{J+1} \right)$. For the products in the first commodity group, consumers will select the product k where $\frac{p_k}{\psi_k} \leq \frac{p_j}{\psi_j}$ for all $j \neq k$. When consumers face the same prices and have homogeneous quality perceptions, they would all be predicted to choose the same product. Changes in a consumer’s income would change the relative proportion of income spent on the commodity group and the essential numeraire. But the income change would not affect her choice of product. Therefore, homotheticity may be particularly problematic in vertically differentiated product categories where observed substitution patterns may be asymmetric between products in different quality tiers. For instance, the cross-elasticity of demand for lower-quality products with respect to premium products’ prices may be higher than the cross-elasticity of demand for higher-quality products with respect to the lower-quality products’ prices (e.g., Blattberg and Wisniewski, 1989; Pauwels, Srinivasan, and Franses, 2007). Similarly, Deaton and Muellbauer (1980b, p. 262) observed a cross-sectional income effect: “richer households systematically tend to buy different qualities than do poorer ones.” Gicheva, Hastings, and Villas-Boas (2010) found a cross-time income effect by showing that lower-income households responded to higher gasoline prices by substituting their grocery purchases towards promotional-priced items, which could be consistent with asymmetric switching patterns if lower-quality items are more likely to

²³Income effects are typically incorporated into demand analyses of high-priced, durable consumption goods like automobiles (e.g., Berry, Levinsohn, and Pakes, 1995).

be promoted. Similarly, Ma, Ailawadi, Gauri, and Grewal (2011) found that households respond to increases in gasoline prices by substituting from national brands to lower-priced brands and to unadvertised own brands supplied by retailers, or “private labels.” These substitution patterns suggest that national brands are normal goods²⁴.

4.1.1 A Non-Homothetic Discrete Choice Model

Given the widespread use of the pure discrete choice models, like logit and probit, we now discuss how to incorporate income effects into these models without losing their empirical tractability. To relax the homotheticity property in the simple re-packaging model with perfect substitutes from section 3.2 above, Deaton and Muellbauer (1980a) and Allenby and Rossi (1991) introduce rotations into the system of linear indifference curves by defining utility implicitly:

$$U(\mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\varepsilon}) = \tilde{U} \left(\sum_j \psi_j(\bar{U}, \boldsymbol{\varepsilon}) x_j, x_{J+1} \right). \quad (49)$$

The marginal utilities in this utility function vary with the level of total attainable utility at the current prices, \bar{U} . If we interpret $\psi_j(\bar{U}, \boldsymbol{\varepsilon})$ as “perceived quality,” then we allow the marginal value of perceived quality to vary with the level of total attainable utility.

To ensure the marginal utilities are positive, Allenby and Rossi (1991) and Allenby, Garratt, and Rossi (2010) use the empirical specification

$$\psi_j(\bar{U}, \boldsymbol{\varepsilon}) = \exp \left(\theta_{j0} - \theta_{j1} \bar{U}(x; \boldsymbol{\theta}) + \varepsilon_j \right)$$

where ε_j is a random utility shock as before. If $\theta_{j1} > 0$, then utility is increasing and concave. The model nests the usual homothetic specification when $\theta_{j1} = 0$ for each product j . To see that the

²⁴Although, Dubé, Hitsch, and Rossi (2017) find highly income-inelastic demand for private label CPGs identified off the large household income shocks during the Great Recession.

parameters θ_{j1} also capture differences in perceived quality, consider the relative marginal utilities:

$$\frac{\psi_k(\bar{U})}{\psi_j(\bar{U})} = \exp\left(\theta_{k0} - \theta_{j0} + (\theta_{j1} - \theta_{k1}) \overline{U(\mathbf{x}; \theta)} + \varepsilon_k - \varepsilon_j\right).$$

The relative perceived quality of product k increases with the level of attainable utility, \bar{U} , so long as $\theta_{k1} < \theta_{j1}$, and so k would be perceived as superior to j . The identification of θ_{k0} comes from the average propensity to purchase product k whereas the identification of θ_{k1} comes from the substitution towards k in response to changes in purchasing power either through budgetary changes to y or through changes in the overall price level.

Consider the case of pure discrete choice: $U(x; \theta) = \left(\sum_{j=1}^{J+1} \psi_j x_j\right) \exp(\psi_{J+1} x_{J+1})$. The consumer chooses between the $j = 1, \dots, J$ products or the $J + 1$ option of allocating the entire budget to the outside good with the following probabilities:

$$Pr(k; \theta) = \frac{\exp(\theta_{k0} - \theta_{k1} \bar{U}^k - \psi_{J+1} p_j)}{1 + \sum_{j=1}^J \exp(\theta_{j0} - \theta_{j1} \bar{U}^j - \psi_{J+1} p_j)}$$

where \bar{U}^k is solved numerically as the solution to the implicit equation

$$\ln(\bar{U}^k) = \theta_{k0} - \theta_{k1} \bar{U}^k - \psi_{J+1} p_j. \quad (50)$$

Maximum likelihood estimation will therefore nest the fixed-point calculation to (50) at each stage of the parameter search.

In their empirical case study of margarine purchases, Allenby and Rossi (1991) find that the demand for generic margarine is considerably more elastic in the price of the leading national brand than vice versa. This finding is consistent with the earlier descriptive findings regarding asymmetric substitution patterns, the key motivating fact for the non-homothetic specification. Allenby, Garratt, and Rossi (2010) project the brand intercepts and utility rotation parameters, θ_{j0} and θ_{j1} respectively, onto advertising to allow the firms' marketing efforts to influence the perceived superiority of their respective brands. In an application to survey data from a choice-based conjoint

survey with a randomized advertising treatment, they find that ads change the substitution patterns in the category by causing consumers to allocate more spending to higher-quality goods.

4.2 Complementary Goods

The determination of demand complementarity figured prominently in the consumption literature (see the survey by Houthakker, 1961). But, the microeconomic literature tackling demand with corner solutions has frequently used additive models that explicitly rule out complementarity and assume products are strict substitutes (Deaton and Muellbauer, 1980b, pages 138-139). For many product categories, such as laundry detergents, ketchups, and refrigerated orange juice, the assumption of strict substitutability seems reasonable for most consumers. However, in other product categories where consumers purchase large assortments of flavors or variants, such as yogurt, carbonated soft drinks, beer and breakfast cereals, complementarity may be an important part of choices. For a shopping basket model that accounts for the wide array of goods, complementarity seems quite plausible between broader commodity groups (e.g. pasta and pasta sauce, or cake mix and frosting).

Economists historically defined complementarity based on the supermodularity of the utility function and the increasing differences in utility associated with joint consumption. Samuelson (1974) provides a comprehensive overview of the arguments against such approaches based on the cardinality of utility. Chambers and Echenique (2009) formally prove that supermodularity is not testable with data on consumption expenditures. Accordingly, most current empirical research defines complementarity based on demand behavior, rather than as a primitive assumption about preferences²⁵.

Perhaps the most widely-cited definition of complementarity comes from Hicks and Allen (1934) using compensated demand:

Definition 1. We say that goods j and k are *complements* if an increase in the price of j leads to a

²⁵An exception is Lee, Kim, and Allenby (2013) who use a traditional definition of complementarity based on the sign of the cross-partial derivative of utility.

decrease in the compensated demand for good k , *substitutes* if an increase in the price of j leads to a decrease in the compensated demand for good k , *independent* if an increase in the price of j has no effect on the compensated demand for good k .

This definition has several advantages including symmetry and the applicability to any number of goods. However, compensated demand is unlikely to be observed in practice. Most empirical research tests for *gross complementarity*, testing for the positivity of the cross-derivatives of Marshallian demands with respect to prices. The linear indifference curves used in most pure discrete choice models eliminates any income effects, making the two definitions equivalent. A recent literature has worked on establishing the conditions under which an empirical test for complementarity is identified with standard consumer purchase data (e.g., Samuelson, 1974; Gentzkow, 2007; Chambers, Echenique, and Shmaya, 2010).

The definition of complementarity based on the cross-price effects on demand can be problematic in the presence of multiple goods. Samuelson (1974, p. 1255) provides the following example:

... sometimes I like tea and cream... I also sometimes take cream with my coffee. Before you agree that cream is therefore a complement to both tea and coffee, I should mention that I take much less cream in my cup of coffee than I do in my cup of tea. Therefore, a reduction in the price of coffee may reduce my demand for cream, which is an odd thing to happen between so-called complements.

To see how this could affect a microeconomic test, consider the model of bivariate utility over a commodity group defined as products in the coffee and cream categories, and an essential numeraire that aggregates expenditures on all other goods (including tea). Even with flexible substitution patterns between coffee and cream, empirical analysis could potentially produce a positive estimate of the cross-price elasticity of demand for cream with respect to the price of coffee if cream is more complementary with tea than with coffee.

On the one hand, this argument highlights the importance of multi-category models, like the

ones we will discuss in section 4.2.2 below, that consider both the intra-category and inter-category patterns of substitution. For instance, one might specify a multivariate utility over all the beverage-related categories and the products within each of the categories. The multi-category model would characterize all the direct and indirect substitution patterns between goods Ogaki (1990, p. 1255). On the other hand, a multi-category model increases the technical and computational burden of demand estimation dramatically. As discussed in Gentzkow (2007, p. 720), the estimated quantities in the narrower, single-category specification should be interpreted as “conditional on the set of alternative goods available in the market.” The corresponding estimates will still be correct for many marketing applications, such as the evaluation of the marginal profits of a pricing or promotional decision. The estimates will be problematic if there is a lot of variation in the composition of the numeraire that, in turn, changes the specification of utility for the commodity group of interest.

Our discussion herein focuses on static theories of complementarity. While several of the models in section 3 allow for complementarity, the literature has been surprisingly silent on the identification strategies for testing complementarity. An exception is Gentzkow (2007), which we discuss in more detail below. A burgeoning literature has also studied the complementarities that arise over time in durable goods markets with inter-dependent demands and indirect network effects. These “platform markets” include such examples as the classic “razors & blades” and “hardware & software” cases (e.g., Ohashi, 2003; Nair, Chintagunta, and Dubé, 2004; Hartmann and Nair, 2010; Lee, 2013; Howell and Allenby, 2017).

4.2.1 Complementarity Between Products Within a Commodity Group

In most marketing models of consumer demand, products within a commodity group are assumed to be substitutes. When a single product in the commodity group is purchased on a typical trip, the perfect substitutes specification is used (see sections 3.2 and 3.3). However, even when multiple products are purchased on a given trip, additive models that imply products are strict substitutes are still used (see for instance the translated CES model in section 3.1.4). Even though some consumer goods products are purchased jointly, they are typically assumed to be consumed separately. There

are of course exceptions. The ability to offer specific bundles of varieties of beverage flavors or beer brands could have a complementary benefit when a consumer is entertaining guests. Outside the CPG domain, Gentzkow (2007) analyzed the potential complementarities of jointly consuming digital and print versions of news.

To incorporate the definition of complementary goods into our demand framework, we begin with a discrete quantity choice model of utility over $j = 1, \dots, J$ goods in a commodity group, where $x_j \in \{0, 1\}$, and a $J + 1$ essential numeraire. The goal consists of testing for complementarity between the goods within a given commodity group. The discussion herein closely follows Gentzkow (2007).

We index all the possible commodity-group bundles the consumer could potentially purchase as $c \in \mathcal{P}(\{1, \dots, J\})$, using $c = 0$ to denote the allocation of her entire budget to the numeraire. The consumer obtains the following choice-specific utility, normalized by u_0

$$u_c = \begin{cases} \sum_{j \in c} (\psi_j - \alpha p_j + \varepsilon_j) + \frac{1}{2} \sum_{j \in c} \sum_{k \in c, k \neq j} \Gamma_{jk}, & \text{if } c \in \mathcal{P}(\{1, \dots, J\}) \\ 0, & \text{if } c = 0 \end{cases} \quad (51)$$

where Γ is symmetric and $\mathcal{P}(\{1, \dots, J\})$ is the power set of the $j = 1, \dots, J$ products. Assume that $\varepsilon \sim N(0, \Sigma)$. To simplify the discussion, suppose that the commodity group comprises only two goods, j and k . The choice probabilities are then

$$\begin{aligned} Pr(j) &= \int_{\{\varepsilon | u_j \geq 0, u_j \geq u_k, u_j \geq u_{jk}\}} dF(\varepsilon) \\ Pr(k) &= \int_{\{\varepsilon | u_k \geq 0, u_k \geq u_j, u_k \geq u_{jk}\}} dF(\varepsilon) \\ Pr(jk) &= \int_{\{\varepsilon | u_{jk} \geq 0, u_{jk} \geq u_j, u_{jk} \geq u_k\}} dF(\varepsilon). \end{aligned}$$

Finally, the expected consumer demand can be computed as follows: $x_j = Pr(j) + Pr(jk)$ and $x_k = Pr(k) + Pr(jk)$.

It is straightforward to show that an empirical test of complementarity between two goods, j and k , reduces to the the sign of the corresponding Γ_{jk} elements of Γ . An increase in the price p_k

has two effects on demand x_j . First, marginal consumers who would not buy the bundle but who were indifferent between buying only j or only k alone will switch to j . At the same time, however, marginal consumers who would not buy only j or only k , and who are indifferent between buying the bundle or not, will switch to non-purchase. More formally,

$$\frac{\partial x_j}{\partial p_k} = \frac{\partial Pr(j)}{\partial p_k} + \frac{\partial Pr(jk)}{\partial p_k} = \int_{\{\varepsilon|u_j=u_k, u_k \geq 0, -\Gamma_{kj} \geq u_j\}} dF(\varepsilon) - \int_{\{\varepsilon|u_j+u_k=-\Gamma_{jk}, u_j \leq 0, u_k \leq 0\}} dF(\varepsilon).$$

We can see that our test for complementarity is determined by the sign of Γ_{jk}

$$\begin{aligned} \Gamma > 0 &\Rightarrow \frac{\partial x_j}{\partial p_k} < 0 \text{ and } j \text{ and } k \text{ are complements} \\ \Gamma = 0 &\Rightarrow \frac{\partial x_j}{\partial p_k} = 0 \text{ and } j \text{ and } k \text{ are independent} \\ \Gamma < 0 &\Rightarrow \frac{\partial x_j}{\partial p_k} > 0 \text{ and } j \text{ and } k \text{ are substitutes.} \end{aligned}$$

Gentzkow (2007) provides a practical discussion of the identification challenges associated with Γ_{jk} , even for this stylized discrete choice demand system. At first glance, (51) looks like a standard discrete choice model where each of the possible permutations of products has been modeled as a separate choice²⁶. But, the correlated error structure in ε plays an important role in the identification of the complementarity. The key moment for the identification of Γ is the incidence of joint purchase of products j and k , $Pr(jk)$. But, high $Pr(jk)$ could arise either through a high value of Γ_{jk} or a high value of $cov(\varepsilon_j, \varepsilon_k)$. A restricted covariance structure like logit, which sets $cov(\varepsilon_j, \varepsilon_k) = 0$ will be forced to attribute a high $Pr(jk)$ to complementarity.

An ideal instrument for testing complementarity would be an exclusion restriction. Consider for instance a variable z_j that shifts the the mean utility for j but does not affect Γ or the mean utility of good k . In the CPG context, the access to high-frequency price variation in all the observed products as well as point-of-purchase promotional variables are ideal for this purpose. The identification of Γ could then reflect the extent to which changes in z_j affect demand x_k .

Panel data can also be exploited to identify Γ_{jk} and $cov(\varepsilon_j, \varepsilon_k)$. Following the conventions in

²⁶For instance, Manski and Sherman (1980) and Train, McFadden, and Ben-Akiva (1987) use logit and nested logit specifications that restrict the covariance patterns in ε .

the literature allowing for persistent, between-consumer heterogeneity, we could let ε be persistent, consumer-specific “random effects.” We could then also include i.i.d. shocks that vary across time and product to explain within-consumer switches in behavior. If joint purchase reflects $cov(\varepsilon_j, \varepsilon_k)$, we would expect to see some consumers frequently buying both and other consumers frequently buying neither. But, conditional on a consumer’s average propensity to purchase either good, the cross-time variation in choices should be uncorrelated. However, if joint purchase reflects Γ_{jk} , we would then expect more correlation over time whereby a consumer would either purchase both goods or neither, but would seldom purchase only one of the two.

4.2.2 Complementarity Between Commodity Groups (multi-category models)

In the analysis of CPG data, most of the emphasis on complementarity has been between product categories, where products within a commodity group are perceived as substitutes but different commodity groups may be perceived as complements. Typically, such cross-category models have been specified using probabilistic choice models without a microeconomic foundation, and that allow for correlated errors either in the random utility shocks (e.g., Manchanda, Ansari, and Gupta, 1999; Chib, Seetharaman, and Strijnev, 2002) or in the persistent heteroskedastic shocks associated with random coefficients (e.g., Ainslie and Rossi, 1998; Erdem, 1998). For an overview of these models, see the discussion in Seetharaman, Chib, Ainslie, Boatright, Chan, Gupta, Mehta, Rao, and Strijnev (2005). The lack of a microfoundation complicates the ability to assign substantive interpretations of model parameters. For instance, the identification discussion in the previous section clarifies the fundamental distinction between correlated tastes (as in these multi-category probabilistic models) and true product complementarity.

At least since Song and Chintagunta (2007), the empirical literature has used microfounded demand systems to accommodate complementarity and substitutability in the analysis of the composition of household shopping baskets spanning many categories during a shopping trip. Conceptually, it is straightforward to extend the general, non-additive frameworks in section 3 to many commodity groups. For instance, Bhat, Castro, and Pinjari (2015) introduce potential complemen-

tarity into the translated CES specification (see section 3.1.4) by relaxing additivity and allowing for interaction effects²⁷. In their study of the pro-competitive effects of multiproduct grocery stores, Thomassen, Seiler, Smith, and Schiraldi (2017) use a quadratic utility model that allows for gross complementarity²⁸. Mehta (2015) uses the indirect translog utility approximation to derive a multi-category model that allows for complementarities.

The direct application of the models in section 3 and the extensions just discussed is limited by the the escalation in parameters and the dimension of numerical integration, both of which grow with the number of products studied. Typically, researchers have either focused their analysis on a small set of product alternatives within a commodity group²⁹ or have focused their analysis on aggregate expenditure behavior across categories, collapsing each category into an aggregated composite good³⁰. As we discuss in the next subsection, additional restrictions on preferences have been required to accommodate product-level demand analysis across categories.

Example: Perfect Substitutes within a Commodity Group Suppose the consumer makes purchase decisions across $m = 1, \dots, M$ commodity groups, each containing $j = 1, \dots, J_m$ products. The consumer has a weakly separable, multivariate utility function over each of the M commodity groups and an $M + 1$ essential numeraire good with price $p_{M+1} = 1$. Within each category, the consumer has perfect substitutes sub-utility over the products, giving consumer utility

$$\tilde{U} \left(\sum_{j=1}^{J_1} \psi_{1j} x_{1j}, \dots, \sum_{j=1}^{J_M} \psi_{Mj} x_{Mj}, \psi_{M+1} x_{M+1} \right) \quad (52)$$

²⁷This specification does not ensure that global regularity is satisfied, which could limit the ability to conduct counterfactual predictions with the model.

²⁸Empirically, they find that positive cross-price elasticities between grocery categories within a store are driven more by shopping costs associated with store choice than by intrinsic complementarities based on substitution patterns between categories.

²⁹For instance, Wales and Woodland (1983) allow for $J = 3$ alternatives of meat: beef, lamb and other meat, and Kim, Allenby, and Rossi (2002) allow for $J = 6$ brand alternatives of yogurt.

³⁰For instance, Kao, fei Lee, and Pitt (2001) look at expenditures across $J = 7$ food commodity groups, Mehta (2015) looks at trip-level expenditures across $J = 4$ supermarket categories, and Thomassen, Seiler, Smith, and Schiraldi (2017) look at trip-level expenditures across $J = 8$ supermarket product categories.

and budget constraint

$$\sum_{m=1}^M \sum_{j=1}^{J_m} p_{mj} x_{mj} + x_{M+1} \leq y.$$

The utility function is a generalization of the discrete choice specification in section 3.2 to many commodity groups. At most, one product will be chosen in each of the M commodity groups. As before, $\psi_{mj} \geq 0$ and $\tilde{U}(\cdot)$ is continuously-differentiable, quasi-concave and increasing function in each of its arguments. We also assume additional regularity conditions to ensure that an interior quantity of the essential numeraire is always purchased, $x_{M+1}^* > 0$. This approach with perfect substitutes within a category has been used in several studies (e.g., Song and Chintagunta, 2007; Mehta, 2007; Lee and Allenby, 2009; Mehta and Ma, 2012). Most of the differences across studies are based on the assumptions regarding the multivariate utility function $\tilde{U}(x)$.

Lee and Allenby (2009) use a primal approach that specifies a quadratic utility over the commodity groups

$$\tilde{U}(u(x_1; \psi_1), \dots, u(x_M; \psi_M), \psi_{M+1} x_{M+1}) = \sum_{m=1}^{M+1} \beta_{m0} u(x_m; \psi_m) - \frac{1}{2} \sum_{m=1}^{M+1} \sum_{n=1}^{M+1} \beta_{mn} u(x_m; \psi_m) u(x_n; \psi_n) \quad (53)$$

where

$$u(x_m; \psi_m) = \sum_{j=1}^{J_m} \psi_{mj} x_{mj}$$

and

$$\psi_{mj} = \exp(\bar{\psi}_{mj} + \varepsilon_{mj}).$$

where $\bar{\psi}_{M+1} = 0$, we normalize β_{10} , and we assume symmetry such that $\beta_{mn} = \beta_{nm}$ for $m, n = 1, \dots, M+1$. The KKT conditions associated with the maximization of the utility function (53) are now as follows:

$$\tilde{\varepsilon}_{mj} = h_{mj}(\mathbf{x}^*; \psi), \text{ if } x_{mj}^* > 0$$

$$\tilde{\varepsilon}_{mj} \leq h_{mj}(\mathbf{x}^*; \psi), \text{ if } x_{mj}^* = 0$$

where $\tilde{\varepsilon}_{mj} = \varepsilon_{mj} - \varepsilon_{M+1}$ and $h_{mj}(\mathbf{x}^*; \psi) = -\ln\left(\frac{\bar{\psi}_{mj}}{p_{mj}}(\beta_{m0} - \sum_{n=1}^M \beta_{mn}u(x_n^*, \psi_n))\right)$. Lee and Allenby (2009) do not impose additional parameter restrictions to ensure the utility function is quasi-concave, a sufficient condition for the coherency of the likelihood function. Instead, they set the likelihood deterministically to zero at any support point where either the marginal utilities are negative or the utility function fails quasi-concavity³¹. While this approach may ensure coherency, it will not ensure that global regularity is satisfied, which could limit the ability to conduct counterfactual predictions with the model.

While products in the same commodity group are assumed to be perfect substitutes, the utility function (53) allows for gross complementarity between a pair of commodity groups, m and n , through the sign of the parameter β_{mn} . In their empirical application, they find gross complementarity between laundry detergents and fabric softeners, which conforms with their intuition. All other pairs of categories studied are found to be substitutes.

Song and Chintagunta (2007), Mehta (2007) and Mehta and Ma (2012) use a dual approach that specifies a translog approximation of the indirect utility function. For simplicity of presentation, we use the homothetic translog specification from Song and Chintagunta (2007)³²

$$V(\mathbf{p}, y; \theta, \varepsilon) = \ln(y) - \sum_{m=1}^{M+1} \theta_m \ln\left(\frac{p_{mj_m}}{\psi_{mj_m}}\right) + \frac{1}{2} \sum_{m=1}^{M+1} \sum_{n=1}^{M+1} \theta_{mn} \ln\left(\frac{p_{mj_m}}{\psi_{mj_m}}\right) \ln\left(\frac{p_{nj_n}}{\psi_{nj_n}}\right) + \sum_{m=1}^{M+1} \varepsilon_{j_m} \ln\left(\frac{p_{mj_m}}{\psi_{mj_m}}\right) \quad (54)$$

where for each commodity group m , product j_m satisfies $\frac{\psi_{mj_m}}{p_{mj_m}} \geq \frac{\psi_{mj}}{p_{mj}}$, $\forall j \neq j_m$. To ensure the coherency of the model, the following parameter restrictions are imposed:

$$\sum_m \theta_m = 1, \theta_{mn} = \theta_{nm}, \forall m, n$$

$$\sum_{m=1}^{M+1} \theta_{mn} = 0, \forall n.$$

³¹The presence of indicator functions in the likelihood create discontinuities that could be problematic for maximum likelihood estimation. The authors avoid this problem by using a Bayesian estimator that does not rely on the score of the likelihood.

³²Mehta and Ma (2012) use a non-homothetic translog approximation which generates a more complicated expenditure share system, but which allows for more flexible income effects.

Applying Roy's identity, we derive the following conditional expenditure shares

$$s_{mj_m}(\mathbf{p}, y; \boldsymbol{\theta}, \boldsymbol{\varepsilon}) = \theta_m - \sum_{n=1}^{M+1} \theta_{mn} \ln \left(\frac{p_{mj_m}}{\Psi_{mj_m}} \right). \quad (55)$$

Since the homothetic translog approximation in (54) eliminates income effects from the expenditure shares, a test for complementarity between a pair of categories m and n amounts to testing the sign of θ_{mn} . In particular, conditional on the chosen products in categories m and n , j_m and j_n respectively, complementarity is identified off the changes in s_{mj_m} due to the quality-adjusted price, $\frac{p_{nj_n}}{\Psi_{nj_n}}$. Hence, several factors can potentially serve as instruments to test complementarity. Changes in the price p_{nj_n} is an obvious source. In addition, if the perceived quality Ψ_{nj_n} is projected onto observable characteristics of product j_n , then independent characteristic variation in product j_n can also be used to identify the complementarity. The switching conditions will be important to account for variation in the identity of the optimal product j_n .

A limitation of this specification is that any complementarity only affects the intensive quantity margin and does not affect the extensive brand choice and purchase incidence margins. Song and Chintagunta (2007) do not detect evidence of complementarities in their empirical application, which may be an artifact of the restricted way in which complementarity enters the model. Mehta and Ma (2012) use a non-homothetic translog approximation that allows for complementarity in purchase incidence as well as the expenditure shares. In their empirical application, they find strong complementarities between the pasta and pasta sauces categories. These findings suggest that the retailer should be coordinating prices across the two categories and synchronizing the timing of promotional discounts.

4.3 Discrete Package Sizes and Non-Linear Pricing

In many consumer goods product categories, product quantities are restricted to the available package sizes. For instance, a customer must choose between specific pre-packaged quantities of liquid laundry detergent (e.g., 32 oz, 64 oz or 128 oz) and cannot purchase an arbitrary, continuous quan-

tity. Early empirical work focused on brand choices, either narrowing the choice set to a specific pack size or collapsing all the pack sizes into a composite brand choice alternative. However, these models ignore the intensive quantity margin and limit the scope of their applicability to decision-making on the supply side.

Firms typically offer an array of pre-packaged sizes as a form of commodity bundling, or “non-linear pricing.” In practice, we expect to see quantity discounts whereby the consumer pays a lower price-per-unit when she buys the larger pack size, consistent with standard second-degree price discrimination (e.g., Varian, 1989; Dolan, 1987). However, several studies have documented cases where firms use quantity-surcharging by raising the price-per-unit on larger pack sizes (e.g., Joo, 2018). The presence of nonlinear pricing introduces several challenges into our neoclassical models of demand (e.g., Howell, Lee, and Allenby, 2016; Reiss and White, 2001). First, any kinks in the pricing schedule will invalidate the use of the Kuhn-Tucker conditions³³. Second, the dual approach that derives demand using Roy’s identity is invalidated by non-linear pricing because Roy’s Identity only holds under a constant marginal price for any given product. An exception is the case of piecewise-linear budget sets (e.g., Hausman, 1985; Howell, Lee, and Allenby, 2016). Third, the price paid per unit of a good depends on a consumer’s endogenous quantity choice, creating a potential self-selection problem in addition to the usual non-negativity problem. To see this potential source of bias, note that the consumer’s budget constraint is $\sum_j p_j(x_j) x_j \leq y$, so the price paid is endogenous as it will depend on unobservable (to the researcher) aspects of the quantity demanded by the consumer.

4.3.1 Expand the Choice Set

One simple and popular modeling approach simply expands the choice set to include all available combinations of brands and pack sizes (e.g., Guadagni and Little, 1983). A separate random utility shock is then added to each choice alternative.

Suppose the consumer makes choices over the $j = 1, \dots, J$ products in a commodity group

³³See Lambrecht, Seim, and Skiera (e.g., 2007) and Yao, Mela, Chiang, and Chen (e.g., 2012) for the analysis of demand under three-part mobile tariffs.

where each product is available in a finite number of pre-packaged sizes, $a \in A_j$. If the consumer has additive preferences and the $j = 1, \dots, J$ products are perfect substitutes, $U(\mathbf{x}, x_{J+1}) = u_1\left(\sum_{j=1}^J \psi_j x_j\right) + u_2(x_{J+1})$, her choice-specific indirect utilities are

$$v_{ja} = u_1(\psi_j x_{ja}) + u_2(y - p_{ja}) + \varepsilon_{ja}, \quad j = 1, \dots, J, \quad a \in A_j \quad (56)$$

$$v_{J+1} = u_2(y) + \varepsilon_{J+1}$$

where $\varepsilon_{aj} \sim i.i.d. EV(0, 1)$, which allows for random perceived utility over the pack size variants of a given product. The probability of choosing pack size a for product k is then

$$Pr(ka; \theta) = \frac{\exp(u_1(\psi_k x_{ka}) + u_2(y - p_{ka}))}{\exp(u_2(y)) + \sum_{\{j=1, \dots, J | a \in A_j\}} \exp(u_1(\psi_j x_{ja}) + u_2(y - p_{ja}))}. \quad (57)$$

To see how one might implement this model in practice, assume the consumer has Cobb-Douglas utility. In this case, $u_1(\psi_k x_{ka}) = \alpha_1 \ln(\bar{\psi}_k) + \alpha_1 \ln(x_{ka})$ where α_1 is the satiation rate over the commodity group. Conceptually, this model could be expanded even further to allow the consumer to purchase bundles of the products to configure all possible quantities that are feasible within the budget constraint.

An important limitation of this specification is that it assigns independent random utility to each pack size of the, otherwise, same product. This assumption would make sense if, for instance, a large 64oz plastic bottle of soda is fundamentally different than a small, 12-oz aluminum can of the same soda. In other settings, we might expect a high correlation in the random utility between two pack-size variants of an otherwise identical product (e.g., 6 pack versus 12 pack of aluminum cans of soda). Specifications that allow for such correlation within-brand, such as nested logit, generalized extreme value or even multinomial probit could work. But, in a setting with many product alternatives, it may not be possible to estimate the full covariance structure between each of the product and size combinations.

In some settings, the temporal separation between choices can be used to simplify the problem.

For instance, Goettler and Clay (2011) and Narayanan, Chintagunta, and Miravete (2007) study consumers' discrete-continuous choices on pricing plan and usage. For instance, providers of mobile data services and voice services typically offer consumers choices between pricing plans that differ in their convexity. In practice, we might not expect the consumer to derive marginal utility from the convexity of the pricing plan, seemingly rendering the pricing plan choice deterministic. But, if the consumer makes a discrete choice between pricing plans in expectation of future usage choices, the expectation errors can be used as econometric uncertainty in the discrete choice between plans.

4.3.2 Models of Pack Size Choice

Allenby, Shively, Yang, and Garratt (2004) use the following Cobb-Douglas utility specification

$$U(\mathbf{x}, x_{J+1}) = \sum_{j=1}^J \sum_{a \in A_j} \alpha_1 \ln(\psi_j x_{ka}) + \alpha_2 \ln(y - p_{ja})$$

where $\psi_j = \exp(\bar{\psi}_j + \varepsilon_j)$ and $\varepsilon_j \sim i.i.d. EV(0, 1)$. In this specification, the utilities of each of the pack sizes for a given product are perfectly correlated. The corresponding, optimal pack size choice for a given product j is deterministic:

$$a_j^* = \max_{a \in A_j} \{ \alpha_1 \ln(x_{ka}) + \alpha_2 \ln(y - p_{ja}) \}$$

and does not depend on ψ_j . The consumer's product choice problem is then the usual maximization across the random utilities of each of the $j = 1, \dots, J$ products corresponding to their respective optimal pack sizes choices. The probability of observing the choice of product k is then

$$Pr(k; \theta, a^*) = \frac{\exp(\alpha_1 \bar{\psi}_k + \alpha_1 \ln(x_{a_k^*}) + \alpha_2 \ln(y - p_{a_k^*}))}{\sum_{j=1}^J \exp(\alpha_1 \bar{\psi}_j + \alpha_1 \ln(x_{a_j^*}) + \alpha_2 \ln(y - p_{a_j^*}))} \quad (58)$$

where a_k is the observed pack size chosen for brand k .

One limitation of the pack size demand specification (58) is that the corresponding likelihood

will not have full support. In particular, variation between pack sizes of a given brand, all else equal, will reject the model. In a panel data version of the model with consumer-specific parameters, within-consumer switching between pack sizes of the same brand over time, all else equal, would reject the model.

Goettler and Clay (2011) and Narayanan, Chintagunta, and Miravete (2007) propose a potential solution to this issue, albeit in a different setting. The inclusion of consumer uncertainty over future quantity needs allows for random variation in pack sizes.

5 Moving Beyond the Basic Neoclassical Framework

5.1 Stock-Piling, Purchase Incidence and Dynamic Behavior

The models discussed so far have treated the timing of purchase as a static consumer decision. According to these models, a consumer allocates her entire budget to the essential numeraire (or outside good), when all of the products' prices exceed their corresponding reservation utilities, as in equation (5) above. Indeed, the literature on price promotions has routinely reported a large increase in sales during promotion weeks (see for instance the literature survey by Blattberg and Neslin (1989) and the empirical generalizations in Blattberg, Briesch, and Fox (1995)).

However, in the case of storable products, consumers may accumulate an inventory and time their purchases strategically based on their expectations about future price changes. An empirical literature has found that price promotions affect both the quantity sold and the timing of purchases through purchase acceleration (e.g., Blattberg, Eppen, and Lieberman, 1981; Scott A. Neslin and Quelch, 1985; Gupta, 1991; Bell, Chiang, and Padmanabhan, 1999). This work estimates that purchase acceleration accounts for between 14 and 50 percent of the promotion effect on quantities sold. Purchase acceleration could simply reflect an increase in consumption. However, more recent work finds that the purchase acceleration could reflect strategic timing based on price expectations. Pesendorfer (2002) finds that while the quantity of ketchup sold is generally higher during periods of low prices, the level depends on past prices. Hendel and Nevo (2006b) find that the magnitude

of the total sales response to a price discount in laundry detergent is moderated by the time since the last price discount. The quantity sold increases by a factor of 4.7 if there was not a sale in the previous week, but only by a factor of 2.0 if there was a sale in the previous week. Using household panel data, Hendel and Nevo (2003) also detect a post-promotion dip in sales levels. Looking across 24 CPG categories, Hendel and Nevo (2006b) find that households pay 12.7% less than if they paid the average posted prices.

Collectively, these findings suggest that households may be timing their purchases strategically to coincide with temporary price discounts. In this case, a static model of demand may overestimate the own-price response. The potential bias on cross-price elasticities is not as clear. In the remainder of this section, we discuss structural approaches to estimate demand with stock-piling and strategic purchase timing based, in part, on price expectations. These models can be used to measure short and long-term price response through counterfactuals.

To the best of our knowledge, Blattberg, Buesing, Peacock, and Sen (1978) were the first to propose a formal economic model of consumer stock-piling based on future price expectations. In the spirit of Becker (1965)'s household production theory, they treat the household as a production unit that maintains a stock of market goods to meet its consumption needs. While the estimation of such a model exceeded the computing power available at that time, Blattberg, Buesing, Peacock, and Sen (1978) find that observable household resources, such as home ownership, and shopping costs, such as car ownership and dual-income status, are strongly associated with deal-proneness. In the macroeconomics literature, Aguiar and Hurst (2007) extend the model to account for the time allocation between "shopping" and "household production" to explain why older consumers tend to pay lower prices (i.e. find the discount periods). In the following sub-sections, we discuss more recent research that has estimated the underlying structure of a model of stock-piling.

5.1.1 Stock-Piling and Exogenous Consumption

Erdem, Imai, and Keane (2003) build on the discrete choice model formulation as in section 3.3. Let $t = 1, \dots, T$ index time periods. At the start of each time period, a consumer has inventory i_t of

a commodity and observes the prices p_t . The consumer can endogenously increase her inventory by purchasing quantities x_{jkt} of each of the j products, where $k \in \{1, \dots, K\}$ indexes the discrete set of available pack sizes. Denote the non-purchase decision as $x_{0t} = 0$. Assume the consumer incurs a shopping cost if she chooses to purchase at least one of the products: $F(\mathbf{x}_t; \tau) = \tau \mathbb{I}_{\{\sum_{j=1}^J x_{jt} > 0\}}$. Her total post-purchase inventory in period t is: $i'_t = i_t + \sum_{j=1}^J x_{jt}$.

After making her purchase decision, the consumer draws an exogenous consumption need that is unobserved to the analyst, $\omega_t \sim F_\omega(\omega)$, and that she consumes from her inventory, i'_t ³⁴. Her total consumption during period t is therefore $c_t = \min(\omega_t, i'_t)$, which is consumed at a constant rate throughout the period³⁵. Assume also that the consumer is indifferent between the brands in her inventory when she consumes the commodity and that she consumes each of them in constant proportion: $c_{jt} = \frac{c_t}{i'_t} i'_{jt}$.

If the consumer runs out of inventory before the end of the period, $\omega_t > c_t$, she incurs a stock-out cost $SC(\omega_t, c_t; \lambda) = \lambda_0 + \lambda_1(\omega_t - c_t)$. The consumer also incurs an inventory carrying cost each period based on the total average inventory held during the period. Her average inventory in period t is $\bar{i}_t = \begin{cases} i'_t - \frac{\omega_t}{2}, & \omega_t \leq i'_t \\ \frac{i'_t}{\omega_t} \frac{i'_t}{2}, & \omega_t > i'_t \end{cases}$. Her total inventory carrying cost is given by $IC(i'_t, \omega_t; \delta) = \delta_0 \bar{i}_t + \delta_1 \bar{i}_t^2$.

Assume the consumer has the following perfect substitutes consumption utility function each period:

$$\begin{aligned} U(i_t, \tilde{i}_t, p_t, \omega_t; \theta) &= \sum_{j=1}^J \psi_j c_{jt} + \psi_{J+1} (y - \sum_{j,k} p_{jkt} x_{jkt} - F(\mathbf{x}_t; \tau) - SC(\omega_t, c_t; \lambda) - IC(i'_t, \omega_t; \delta)) \\ &= \frac{c(i'_t, \omega_t)}{i'_t} \tilde{i}_t + \psi_{J+1} (y - \sum_{j,k} p_{jkt} x_{jkt} - F(\mathbf{x}_t; \tau) - SC(\omega_t, c(i'_t, \omega_t); \lambda) - IC(i'_t, \omega_t; \delta)) \end{aligned} \quad (59)$$

where $\tilde{i}_t \equiv \sum_{j=1}^J \psi_j i'_{jt}$ is the post-purchase quality-adjusted inventory, and where the shopping cost, inventory carrying cost and stock-out cost have all been subsumed into the budget constraint. The

³⁴The consumer therefore makes purchase decisions in anticipation of her future expected consumption needs, as in Dubé (2004).

³⁵Sun, Neslin, and Srinivasan (2003) use a data-based approach that measures the exogenous consumption need as a constant consumption rate, based on the household's observed average quantity purchased.

vector $\theta = (\psi_1, \dots, \psi_{J+1}, \lambda_0, \lambda_1, \delta_0, \delta_1, \tau_0)$ contains all the model's parameters.

The three state variables are summarized by $s_t = (i_t, \tilde{i}_t, p_t)$. The inventory state variables evolve as follows:

$$\begin{aligned} i_t &= \left(i_{t-1} + \sum_{j=1}^J x_{jt} \right) \left(1 - \frac{c_t}{i_t} \right) \\ \tilde{i}_t &= \left(\tilde{i}_{t-1} + \sum_{j=1}^J \psi_j x_{jt} \right) \left(1 - \frac{c_t}{i_t} \right) \end{aligned} .$$

Assume in addition that consumers' price beliefs are known to the analyst and evolve according to the Markov transition density³⁶ $p_{t+1} \sim f_p(p_{t+1}|p_t)$.³⁷ Therefore, the state vector also follows a Markov Process which we denote by the transition density $f_s(s'|s, x_{jk})$.

The consumer's purchase problem is dynamic since she can control her future inventory states with her current purchase decision. Assuming the consumer discounts future utility at a rate $\beta \in (0, 1)$, the value function associated with her purchase decision problem in state s_t is

$$v(s_t, \varepsilon_t) = \max_{j,k} \{ v_{jk}(s_t; \theta) + \varepsilon_{jkt} \} \quad (60)$$

where $\varepsilon_{jkt} \sim i.i.d.EV(0, 1)$ is a stochastic term known to the household at time t but not to the analyst. $v_{jk}(s)$ is the choice-specific value function associated with choosing product j and pack size k in state s

$$v_{jk}(s; \theta) = \int U(s_t, \omega; \theta) f_\omega(\omega) d\omega + \beta \int v(s', \varepsilon) f_s(s'|s, x_{jk}) d(s', \varepsilon). \quad (61)$$

When the taste parameters θ are known, the value functions in (60) and (61) can be solved numerically (see for instance Erdem, Imai, and Keane (2003) for technical details).

Suppose the researcher observes a consumer's sequence of choices, $\hat{\mathbf{x}} = (\hat{x}_1, \dots, \hat{x}_T)$. Conditional on the state, s , the probability that the consumer's optimal choice is product j and pack size

³⁶Typically, a rational expectations assumption is made and the price process, $F_p(p_{t+1}|p_t)$, is estimated in a first stage using the observed price series. An interesting exception is Erdem, Keane, Öncü, and Strebel (2005) who elicit consumers' subjective price beliefs through a consumer survey.

³⁷In four CPG case studies, Liu and Balachander (2014) find that a proportional hazard model for the price process fits the price data better and leads to a better fit of the demand model when used to capture consumers' price expectations.

k has the usual multinomial logit demand form:

$$Pr(x_{jk}|s; \theta) = \frac{\exp(v_{jk}(s; \theta))}{\exp(v_0(s; \theta)) + \sum_{k=1}^J \exp(v_{jk}(s; \theta))}, x_{jk} \in \{x_{11}, \dots, x_{JK}\}. \quad (62)$$

To accommodate the fact that the two inventory state variables, i and \tilde{i} , are not observed, we partition the state as follows: $s = (p, \tilde{s})$ where $\tilde{s} = (i, \tilde{i})$. Since we do not observe the initial values of \tilde{s}_0 , we have a classic initial conditions problem (Heckman, 1981). We resolve the initial conditions problem by assuming there is a true initial state, \tilde{s}_0 , with density $f_s(\tilde{s}_0; \theta)$.

We can now derive the density associated with the consumer's observed sequence of purchase decisions, $\hat{\mathbf{x}}$:

$$f(\hat{\mathbf{x}}; \theta) = \int \left(\prod_{t=1}^T \int \prod_{j,k} Pr(x_{jk}|p_t, \tilde{s}_t, \omega, \tilde{s}_0; \theta)^{\mathbb{I}\{x_{jk}=\hat{x}_t\}} f_\omega(\omega) d\omega \right) f_s(\tilde{s}_0; \theta) d\tilde{s}_0. \quad (63)$$

Consistent estimates of the parameters θ can then be obtained via simulated maximum likelihood.

5.1.2 Stock-Piling and Endogenous Consumption

The model in the previous section assumed an exogenous consumption rate, which implies that any purchase acceleration during a price discount period reflects stock-piling. Sun (2005), Hendel and Nevo (2006a) and Liu and Balachander (2014) allow for endogenous consumption. This important extension allows for two types of response to a temporary price cut. In addition to stock-piling, consumers can potentially increase their consumption of the discounted product. We focus on the formulation in Hendel and Nevo (2006a), which reduces part of the computational burden of Erdem, Imai, and Keane (2003) by splitting the likelihood into a static and a dynamic component.

A key assumption is that consumers only value the brand at the time of purchase so that the optimal consumption decisions are independent of specific brands and depend only on the quantities purchased. During period t , the consumer derives the following consumption utility³⁸

³⁸Hendel and Nevo (2006a) also allow for point-of-purchase promotional advertising, like feature ads and displays, to shift utility in their specification.

$$U(c_t + \omega_t, i_t, \mathbf{p}_t; \theta) = u(c_t + \omega_t; \gamma_c) + \sum_{j=1}^J \sum_{k=1}^K \mathbb{I}_{\{x_{jkt} > 0\}} \{ \Psi_{J+1}(y - p_{jkt}) + \Psi_{jk} \} - C(i_{t+1}; \lambda_c) \quad (64)$$

where as before we index the products by $j = 1, \dots, J$ and the discrete pack sizes available by $k \in \{1, \dots, K\}$. $u(c + \omega; \gamma_c)$ is the consumption utility with taste parameters γ_c and $\omega \sim F_\omega(\omega)$ is a random “consumption need” shock. The start of period inventory is $i_t = i_{t-1} + \sum_j \sum_k x_{jkt-1} - c_{t-1}$. $C(i_{t+1}; \lambda_c)$ is the inventory carrying cost with cost-related parameters λ_c . As before, x denotes a purchase quantity (as opposed to consumption), and Ψ_{J+1} captures the marginal utility of the numeraire good.

The three state variables are summarized by $s_t = (i_t, \mathbf{p}_t, \omega_t)$. Inventory evolves as follows:

$$i_t = i_{t-1} + \sum_{j=1}^J x_{jt-1} - c_{t-1} .$$

In addition, consumers’ form Markovian price expectations: $\mathbf{p}_{t+1} \sim F_p(\mathbf{p}_{t+1} | \mathbf{p}_t)$. Unlike Erdem, Imai, and Keane (2003), the consumption need $\omega \sim F_\omega(\omega)$ is a state variable. The state variables, s_t , follow a Markov Process with transition density $f_s(s' | s, x_{jk})$.

The value function associated with the consumer’s purchase decision problem during period t is

$$v(s_t, \boldsymbol{\varepsilon}_t) = \max_{j,k} \{ v_{jk}(s_t) + \boldsymbol{\varepsilon}_{jkt} \} \quad (65)$$

where s_t is the state in period t , $\boldsymbol{\varepsilon}_{jkt} \sim i.i.d.EV(0, 1)$ is a stochastic term known to the household at time t but not to the analyst, and $v_{jk}(s)$ is the choice-specific value function associated with choosing product j and pack size k in state s

$$v_{jk}(s_t) = \Psi_{J+1}(y - p_{jkt}) + \Psi_{jk} + M(s_t, x_{jk}; \theta_c) \quad (66)$$

where $M(s_t, x_{jk}; \theta_c) = \max_c \{ u(c + \omega_t; \gamma_c) - C(i_{t+1}; \lambda_c) + \beta \int v(s', \boldsymbol{\varepsilon}) f_s(s' | s, x_{jk}, c) f_\boldsymbol{\varepsilon}(\boldsymbol{\varepsilon}) d(s', \boldsymbol{\varepsilon}) \}$ and θ_c are the consumption-related parameters.

Hendel and Nevo (2006a) propose a simplified three-step approach to estimating the model parameters. The value function (65) can be simplified by studying Equation (66), which indicates that consumption is only affected by the quantity purchased, not the specific brand chosen. In a first step, it is straightforward to show that consistent estimates of the brand taste parameters, $\boldsymbol{\psi} = (\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_{J+1})$ can be obtained from the following standard multinomial logit model of brand choice across all brands available in the pack size k :

$$Pr(j|s_t, k; \boldsymbol{\psi}) = \frac{\exp(\boldsymbol{\psi}_{J+1}(y - p_{jkt}) + \boldsymbol{\psi}_{jk})}{\sum_i \exp(\boldsymbol{\psi}_{J+1}(y - p_{ikt}) + \boldsymbol{\psi}_{ik})}.$$

In a second step, define the expected value of the optimal brand choice, conditional on pack size, as follows:

$$\eta_{kt} = \log \left\{ \sum_j \exp(\boldsymbol{\psi}_{J+1}(y - p_{jkt}) + \boldsymbol{\psi}_{jk}) \right\}. \quad (67)$$

Using an idea from Melnikov (2013), assume that η_{t-1} is a sufficient statistic for η_t so that $F(\eta_t|s_{t-1})$ can be summarized by $F(\eta_t|\eta_{t-1})$. The size-specific inclusive values can then be computed with the brand taste parameters, $\boldsymbol{\psi}$, and equation (67) and then used to estimate the distribution $F(\eta_t|\eta_{t-1})$.

In a third step, the quantity choice state can then be defined as $\tilde{s}_t = (i_t, \eta_t, \omega_t)$, which reduces dimensionality by eliminating any brand-specific state variables. The value function associated with the consumer's quantity decision problem can now be written in terms of these size-specific "inclusive value" terms:

$$v(\tilde{s}_t, \boldsymbol{\varepsilon}_t) = \max_{c,k} \{u(c + \boldsymbol{\omega}_t; \boldsymbol{\gamma}_c) - C(i_{t+1}; \boldsymbol{\lambda}_c) + \eta_{kt} + \beta \int v(\tilde{s}', \boldsymbol{\varepsilon}) f_s(\tilde{s}'|\tilde{s}, x_k, c) f_\varepsilon(\boldsymbol{\varepsilon}) d(\tilde{s}', \boldsymbol{\varepsilon})\}. \quad (68)$$

Similarly, the pack-size choice-specific value functions can also be written in terms of these size-

specific “inclusive value” terms:

$$v_k(\tilde{s}_t) = \eta_{kt} + M_k(\tilde{s}_t; \theta_c) \quad (69)$$

where $M_k(\tilde{s}_t; \theta_c) = \max_c \{u(c + \omega_t; \gamma_c) - C(i_{t+1}; \lambda_c) + \beta \int v(\tilde{s}', \varepsilon) f_s(\tilde{s}' | \tilde{s}, x_k, c) f_\varepsilon(\varepsilon) d(\tilde{s}', \varepsilon)\}$. The corresponding optimal pack size choice probabilities are then:

$$Pr(k | \tilde{s}_t; \theta_c) = \frac{\exp(\eta_k + M_k(\tilde{s}_t; \theta_c))}{\sum_i \exp(\eta_i + M_i(\tilde{s}_t; \theta_c))}$$

The density associated with the consumer’s observed sequence of pack size decisions, $\hat{\mathbf{x}}$, is³⁹:

$$f(\hat{\mathbf{x}}; \theta_c) = \int \int \left(\prod_{t=1}^T \int Pr(\hat{x}_t | \tilde{s}_t, \tilde{s}_0; \theta_c) f_\omega(\omega) d\omega \right) f_\omega(\omega) f_s(\tilde{s}_0; \theta_c) d(\omega, \tilde{s}_0). \quad (70)$$

Consistent estimates of the parameters θ_c can then be obtained via simulated maximum likelihood. A limitation of this three-step approach is that it does not allow for persistent, unobserved heterogeneity in tastes.

5.1.3 Empirical Findings with Stock-piling Models

In a case study of household purchases of Ketchup, Erdem, Imai, and Keane (2003) find that the dynamic stock-piling model described above fits the data well in-sample, in particular the timing between purchases. Using simulations based on the estimates, they find that a product’s sales-response to a temporary price cut mostly reflects purchase acceleration and category expansion, as opposed to brand switching. This finding is diametrically opposite to the conventional wisdom that “brand switchers account for a significant portion of the immediate increase volume due to sales promotion” (Blattberg and Neslin, 1989, p. 82). The cross-price elasticities between brands are found to be quite small compared to those from static choice models; although the exact magnitude is sensitive to the specification of the price process representing consumers’ expectations. They

³⁹The initial conditions can be resolved in a similar manner as in Erdem, Imai, and Keane (2003).

find much larger cross-price elasticities in response to permanent price changes and conclude that long-run price elasticities are likely more relevant to policy analysts who want to measure the intensity of competition between brands.

In case studies of Packaged Tuna and Yogurt, Sun (2005) finds that consumption increases with the level of inventory and decreases in the level of promotional uncertainty. While promotions do lead to brand-switching, she also finds that they increase consumption. A model that assumes exogenous consumption over-estimates the extent of brand switching.

In a case study of Laundry Detergents, Hendel and Nevo (2006a) focus on the long-run price elasticities by measuring the effects of permanent price changes. They find that a static model generates 30% larger price elasticities than the dynamic model. They also find that the static model underestimates cross-price elasticities. Some of the cross-price elasticities in the dynamic model are more than 20 times larger than those from the static model. Finally, the static model overestimates the degree of substitution to the outside good by 200%. Seiler (2013) builds on Hendel and Nevo (2006a)'s specification by allowing consumers with imperfect price information to search each period before making a purchase decision. In a case study of laundry detergent purchases, Seiler (2013)'s parameter estimates imply that 70% of consumers do not search each period. This finding highlights the importance of merchandizing efforts, such as in-store displays, to help consumers discover low prices. In addition, by using deeper price discounts, a firm can induce consumers to engage in more price search which can increase total category sales. This increase in search offsets traditional concerns about inter-temporal cannibalization due to strategic purchase timing.

5.2 The Endogeneity of Marketing Variables

The frameworks discussed thus far focus entirely on the demand side of the market. However, many of the most critical demand-shifting variables at the point of purchase consist of marketing mix variables such as prices and promotions, including merchandizing activities like temporary discounts, in-aisle displays and feature advertising. If these marketing variables are set strategi-

cally by firms with more consumer information than the researcher, any resulting correlation with unobserved components of demand could impact the consistency of the likelihood-based estimates discussed thus far. In fact, one of the dominant themes in the empirical literature on aggregate demand estimation consists of the resolution of potential endogeneity of supply-side variables (e.g., Berry, 1994; Berry, Levinsohn, and Pakes, 1995). While most of the literature has focused on obtaining consistent demand estimates in the presence of endogenous prices, bias could also arise from the endogeneity of advertising, promotions and other marketing variables⁴⁰. Surprisingly little attention has been paid to the potential endogeneity of marketing variables in the estimation of individual consumer level demand. In the remainder of this section, we focus on the endogeneity of prices even though many of the key themes would readily extend to other endogenous demand-shifting variables⁴¹.

Suppose a sample of $i = 1, \dots, N$ consumers each makes a discrete choice among $j = 1, \dots, J$ product alternatives and a $J + 1$ “no purchase” alternative. Each consumer is assumed to obtain the following conditional indirect utility from choice j :

$$V_{ij} = \psi_j - \alpha p_{ij} + \varepsilon_{ij}$$

$$V_{iJ+1} = \varepsilon_{i,J+1}$$

where $\varepsilon_i \sim i.i.d. F(\varepsilon)$ and p_{ij} is the price charged to consumer i for alternative j . Demand estimation is typically carried out by maximizing the corresponding likelihood function:

$$L = \prod_i \prod_j Pr(j; \theta)^{y_{ij}} \quad (71)$$

where $Pr(j; \theta) = Pr(V_{ij} \geq V_{ik}, \forall k \neq j)$ and $y = (y_{i1}, \dots, y_{iJ+1})$ indicates which of the $j = 1, \dots, J +$

⁴⁰For instance, Manchanda, Rossi, and Chintagunta (2004) address endogenous detailing levels across physicians.

⁴¹In the empirical consumption literature, the focus has been more on the endogeneity of household incomes than the endogeneity of prices (see for instance Blundell, Pashardes, and Weber, 1993). Since the analysis is typically at the broad commodity group level (e.g., food), the concern is that household budget shares are determined simultaneously with consumption quantities.

1 product was chosen by consumer i .

If $\text{cov}(p_{ij}, \varepsilon_{ij}) \neq 0$ then the maximum likelihood estimator θ^{MLE} based on (71) may be inconsistent since the likelihood omits information about ε . In general, endogeneity can arise in three ways (see Wooldridge, 2002, for example):

1. Simultaneity: Firms observe and condition on ε_i when they set their prices
2. Self-Selection: Certain types of consumers systematically find the lowest prices
3. Measurement Error: The researcher observes a noisy estimate of true prices, $\tilde{p}_{ij} : \tilde{p}_{ij} = p_{ij} + \eta_{ij}$.

Most of the emphasis has been on simultaneity bias whereby endogeneity arises because of the strategic pricing decisions by the firms. Measurement error is not typically discussed in the demand estimation literature. However, many databases contain time-aggregated average prices rather than the actual point-of-purchase price, which could lead to classical measurement error. To the best of our knowledge, a satisfactory solution has yet to be developed for demand estimation with this type of measurement error. In many marketing settings, endogeneity bias could also arise from the self-selection of consumers into specific marketing conditions based on unobserved (to the researcher) aspects of their tastes. For instance, unobserved marketing promotions like coupons could introduce both measurement error and selection bias if certain types of consumers are systematically more likely to find/have a coupon and use it (Erdem, Keane, and Sun, 1999). Similarly, Howell, Lee, and Allenby (2016) propose an approach to resolve the price self-selection bias associated with consumers choosing between non-linear pricing contracts based on observable (to the researcher) aspects of their total consumption needs. If consumers face incomplete information about the choice set, then selection could arise from price search and the formation of consumers' consideration sets (e.g., Honka, 2014). The topics of consumer search and the formation of consideration sets are discussed in more detail in Chapters in this volume on branding and on search. Finally, the potential self-selection of consumers into discount and regular prices based on their unobserved (to the researcher) potential stock-piling behavior during promotional periods in an-

anticipation of future price increases could also bias preference estimates (e.g., Erdem, Imai, and Keane, 2003; Hendel and Nevo, 2006a).

For the remainder of this discussion, we will focus on price endogeneity associated with the simultaneity bias. Suppose that $j = 1, \dots, J$ consumer goods in a product category are sold in $t = 1, \dots, T$ static, spot markets by single-product firms playing a Bertrand-Nash pricing game. Typically, a market is a store-week since stores tend to set their prices at a weekly frequency and most categories in the store are “captured markets” in the sense that consumers likely to not base their store choices on each of the tens of thousands of prices charged across the products carried in a typical supermarket. On the demand side, consumers make choices in each market t to maximize their choice-specific utility

$$V_{ijt} = v_j(\mathbf{w}_t, \mathbf{p}_t; \theta) + \xi_{jt} + \varepsilon_{ijt} \quad (72)$$

$$V_{iJ+1t} = \varepsilon_{i,J+1t}$$

where we distinguish between the exogenous point-of-purchase utility shifters, w_t , and the prices, p_t . In addition, we now specify a composite error term consisting of the idiosyncratic utility shock, $\varepsilon_{ijt} \sim i.i.d. EV(0, 1)$, and the common shock, $\xi_{jt} \sim i.i.d. F_\xi(\xi)$, to control for potential product- j specific characteristics that are observed to the firms when they set prices, but not to the researcher (Berry, 1994). Consumers have corresponding choice probabilities, $Pr(j; \theta | \mathbf{w}_t, \mathbf{p}_t, \xi_t)$ for each of the j alternatives including the $J + 1$ no-purchase alternative. Price endogeneity arises when the firms conditions on ξ when setting its prices and $cov(\mathbf{p}_t, \xi_t) \neq 0$.

A consistent and efficient estimator can be constructed by maximizing the following likelihood

$$L(\theta | \hat{\mathbf{x}}) = \prod_t \int \cdots \int \prod_i \prod_j Pr(j; \theta | \mathbf{w}_t, \mathbf{p}_t, \xi) y_{ijt} f_p(\mathbf{p}_t | \xi) f_\xi(\xi) d\xi_1 \dots d\xi_J.$$

In practice, the form of the likelihood of prices may not be known and ad hoc assumptions about $f_p(\mathbf{p} | \xi)$ could lead to additional specification error concerns. We now discuss the trade-offs between full-information and limited information approaches.

5.2.1 Incorporating the Supply Side: A Structural Approach

An efficient “full-information” solution to the price endogeneity bias consists of modeling the data-generating process for prices and deriving the density $f_p(\mathbf{p}|\xi)$ *structurally*. Since consumer goods are typically sold in a competitive environment, this approach requires specifying the structural form of the pricing game played by the various suppliers. The joint density of prices is then induced by the equilibrium in the game (e.g., Yang, Chen, and Allenby, 2003; Draganska and Jain, 2004; Villas-Boas and Zhao, 2005).

On the supply side of the model in equation (72), assume the J firms play a static, Bertrand-Nash game for which the prices each period satisfy the following necessary conditions for profit maximization:

$$Pr(j; \theta | \mathbf{w}_t, \mathbf{p}_t, \xi_t) + (p_{jt} - c_{jt}) \frac{\partial Pr(j; \theta | p_t, \xi_t)}{\partial p_{jt}} = 0 \quad (73)$$

where $c_{jt} = b_{jt}\gamma + \eta_{jt}$ is firm j 's marginal cost in market t , b_{jt} are observable cost-shifters, like factor prices, γ are the factor weights and $\eta_t \sim i.i.d. F(\eta)$ is a vector of cost shocks that are unobserved to the researcher. We use the static Nash concept as an example. Alternative modes of conduct (including non-optimal behavior) could easily be accommodated instead. In general, these first order conditions (73) will create covariance between prices and demand shocks, $cov(\mathbf{p}_t, \xi_t) \neq 0$.

As long as the system of first-order conditions, (73), generates a unique vector of equilibrium prices, we can then derive the density of prices $f_p(\mathbf{p}|\xi_t, \mathbf{c}_t, \mathbf{w}_t) = f(\eta_t|\xi_t) |J_{\eta \rightarrow p}|$ where $J_{\eta \rightarrow p}$ is the Jacobian of the transformation from η to prices. A consistent and efficient estimate of the parameters $\Theta' = (\theta', \gamma')$ can then be obtained by maximizing the likelihood function⁴²

$$L(\Theta) = \prod_t \int \int \prod_i \prod_j Pr(j; \theta | \mathbf{w}_t, \mathbf{p}_t, \xi_t)^{y_{ijt}} f_p(\mathbf{p}_t | \xi_t, \mathbf{c}_t, \mathbf{w}_t) f(\xi_t) d\eta d\xi. \quad (74)$$

Two key concerns with this approach are as follows. First, in many settings, pricing conduct may be more sophisticated than the single-product, static Bertrand Nash setting characterized by

⁴²Yang, Chen, and Allenby (2003) propose an alternative Bayesian MCMC estimator.

(73). Mis-specification of the pricing conduct would lead to a mis-specification of the density $f_p(\mathbf{p}|\xi_t)$, which could lead to bias in the parameter estimates. Yang, Chen, and Allenby (2003) resolve this problem by testing between several different forms of static, pricing conduct. An advantage of their Bayesian estimator is the ability to re-cast the conduct test as a Bayesian decision theory problem of model selection. Villas-Boas and Zhao (2005) incorporate conduct parameters into the system of first-order necessary conditions (73), where specific values of the conduct parameter nest various well-known pricing games.

In addition to the conduct specification, even if we can assume existence of a price equilibrium, for our simple static Bertrand-Nash pricing game it is difficult to prove uniqueness to the system of first-order necessary conditions (73) for most demand specifications, $Pr(j; \theta | \mathbf{w}_t, \mathbf{p}_t, \xi_t)$. This non-uniqueness problem translates into a coherency problem for the maximum likelihood estimator based on 74. The multiplicity problem would likely be exacerbated in more sophisticated pricing games involving dynamic conduct, multi-product firms and channel interactions. Berry, Levinsohn, and Pakes (1995) avoid this problem by using a less efficient GMM estimation approach that does not require computing the Jacobian term $J_{\eta \rightarrow p}$. Another potential direction for future research might be to recast (71) as an incomplete model and to use partial identification for inference on the supply and demand parameters (e.g., Tamer, 2010).

A more practical concern is the availability of exogenous variables, b_{jt} , that shift prices but are plausibly excluded from demand. Factor prices and other cost-related factors from the supply side may be available. In the absence of any exclusion restrictions, identification of the demand parameters will then rely on the assumed structure of $f_p(\mathbf{p}|\xi) f(\xi)$.

The full-information approaches have thus far produced mixed evidence on the endogeneity bias in the demand parameters in a small set of empirical case studies. Draganska and Jain (2004) and Villas-Boas and Zhao (2005) find substantial bias, especially in the price coefficient α . However, in a case study of light beer purchases, Yang, Chen, and Allenby (2003) find that the endogeneity bias may be an artifact of omitted heterogeneity in the demand specification. Once they allow for unobserved demand heterogeneity, they obtain comparable demand estimates regardless

of whether they incorporate supply-side information into the likelihood. Interestingly, in a study of targeted detailing to physicians, Manchanda, Rossi, and Chintagunta (2004) find that incorporating the supply side not only resolves asymptotic bias in the estimates of demand parameters, they also find a substantial improvement in efficiency⁴³.

5.2.2 Incorporating the Supply Side: A Reduced-Form Approach

As explained in the previous section, the combination of potential specification error and a potential multiplicity of equilibria are serious disadvantages to full-information approaches. In the literature, several studies have proposed less efficient limited-information approaches that are more agnostic about the exact data-generation process on the supply side.

Villas-Boas and Winer (1999) use a more agnostic approach that is reminiscent of two-stage least squares estimators in the linear models setting. Rather than specify the structural form of the pricing game on the supply side, they instead model the reduced form of the equilibrium prices

$$p_{jt} = W_j(w_t, b_t; \lambda) + \zeta_{jt} \quad (75)$$

where b_t are again exogenous, price-shifters that are excluded from the demand side, and ζ_t is a random price shock such that $(\xi'_t, \zeta'_t) \sim F(\xi_t, \zeta_t)$ and b_t are independent of (ξ'_t, ζ'_t) . It is straightforward to derive $f(\mathbf{p}|\xi_t, b_t, \mathbf{w}_t) = f(\zeta_t|\xi_t)$ since the linearity obviates the need to compute a Jacobian. A consistent “limited information” estimate of the parameters $\Theta' = (\theta', \lambda)'$ can then be obtained by substituting this density into the likelihood function (74). While this approach does not require specifying pricing conduct, unlike a two-stage least squares estimator, linearity is not an innocuous assumption. Any specification error in the ad hoc “reduced form” will potentially bias the demand estimates. For instance, the first-order necessary conditions characterizing equilibrium prices in (73) would not likely reduce to a specification in which the endogenous component of prices is an additive, Gaussian shock. Conley, Hansen, McCulloch, and Rossi (2008) resolve

⁴³Manchanda, Rossi, and Chintagunta (2004) address a much more sophisticated form of endogeneity bias whereby the detailing levels are coordinated with the firm’s posterior beliefs about a physician’s response coefficients, as opposed to an additive error component as in the cases discussed above.

this problem by using a semi-parametric, mixture-of-Normals approximation of the density over (ξ'_t, ζ'_t) .

A separate stream of work has developed instrumental variables methods to handle the endogeneity of prices. Chintagunta, Dubé, and Goh (2005) conduct a case study of product categories in which, each store-week, a large number of purchases are observed for each product alternative. On the demand side, they can then directly estimate the weekly mean utilities as “fixed effects”

$$V_{ijt} = v_j(\mathbf{w}_t, \mathbf{p}_t; \theta) + \xi_{jt} + \varepsilon_{ijt}$$

$$V_{iJ+1t} = \varepsilon_{i,J+1t}$$

without needing to model the supply side. Using standard maximum likelihood estimation techniques, they estimate the full set of brand-week effects $\{\psi_{jt}\}_{j,t}$ in a first stage⁴⁴. Following Nevo (2001)’s approach for aggregate data, the mean responses to marketing variables are obtained in a second stage minimum distance procedure that projects the brand-week effects onto the product attributes, \mathbf{x}_t and \mathbf{p}_t

$$\hat{\psi}_{jt} = v_j(\mathbf{w}_t, \mathbf{p}_t; \theta) + \xi_{jt} \tag{76}$$

using instrumental variables, $(\mathbf{w}_t, \mathbf{b}_t)$ to correct for the potential endogeneity of prices. Unlike Villas-Boas and Winer (1999), the linearity in 75 does not affect the consistency of the demand estimates. Even after controlling for persistent, unobserved consumer taste heterogeneity, Chintagunta, Dubé, and Goh (2005) find strong evidence of endogeneity bias in both the levels of the response parameters and in the degree of heterogeneity. A limitation of this approach is that any small sample bias in the brand-week effects will potentially lead to inconsistent estimates.

In related work, Goolsbee and Petrin (2004) and Chintagunta and Dubé (2005) use an alternative approach that obtains exact estimates of the mean brand-week utilities by combining the individual purchase data with store-level data on aggregate sales. Following Berry, Levinsohn, and Pakes (1995) (BLP), the weekly, mean brand-week utilities are inverted out of the observed weekly

⁴⁴Their estimator allows for unobserved heterogeneity in consumers’ responses to marketing variables.

marke share data, \mathbf{s}_t

$$\boldsymbol{\psi}_t = Pr^{-1}(\mathbf{s}_t) \quad (77)$$

where $Pr^{-1}(\mathbf{s}_t)$ is the inverse of the system of predicted market shares corresponding to the demand model⁴⁵. These mean utilities are then substituted into the first stage for demand estimation⁴⁶. In a second stage, the mean response parameters are again obtained using the projection (76) and instrumental variables to correct for the endogeneity of prices.

When aggregate market share data are unavailable, Petrin and Train (2010) propose an alternative “control function” approach. On the supply side, prices are again specified in reduced form as in (75). On the demand side, consumers make choices in each market t to maximize their choice-specific utility

$$V_{ijt} = v_j(\mathbf{w}_t, \mathbf{p}_t; \boldsymbol{\theta}) + \varepsilon_{ijt}$$

$$V_{iJ+1t} = \varepsilon_{i,J+1t}$$

where the utility shocks to the $j = 1, \dots, J$ products can be decomposed as follows:

$$\varepsilon_{ijt} = \varepsilon_{ijt}^1 + \varepsilon_{ijt}^2$$

where $(\varepsilon_{ijt}^1, \zeta_{ijt}) \sim N(0, \Sigma)$ and $\varepsilon_{ijt}^2 \sim i.i.d. F(\varepsilon)$. We can then re-write the choice-specific utility as:

$$V_{ijt} = v_j(\mathbf{w}_t, \mathbf{p}_t; \boldsymbol{\theta}) + \lambda \zeta_{ijt} + \sigma \eta_{jt} + \varepsilon_{ijt}^2, j = 1, \dots, J \quad (78)$$

where $\eta_{jt} \sim N(0, 1)$. Estimation is then conducted in two steps. The first stage consists of the price regression based on equation (75). The second stage consists of estimating the choice probabilities corresponding to (78) using the control function, $\lambda \zeta$ for alternatives $j = 1, \dots, J$ with parameter λ to be estimated. In an application to household choices between satellite and cable television content suppliers, Petrin and Train (2010) find that the control function in (78) generates compa-

⁴⁵See Berry (1994) and Berry, Gandhi, and Haile (2013) for the necessary and sufficient conditions required for the demand system to be invertible.

⁴⁶Chintagunta and Dubé (2005) estimate the parameters characterizing unobserved heterogeneity in this first stage

rable demand estimates to those obtained using the more computationally and data-intensive BLP approach based on (77).

5.3 Behavioral Economics

The literature on behavioral economics has created an emerging area for microeconomic models of demand. This research typically starts with surprising or puzzling moments in the data that would be difficult to fit using the standard neoclassical models. In this section, we look at two specific topics: the fungibility of income and social preferences. For a broader discussion of structural models of behavioral economics, see DellaVigna (2017). The pursuit of ways to incorporate more findings from the behavioral economics literature into quantitative models of demand seems like a fertile area for future research⁴⁷.

5.3.1 The Fungibility of Income

Building on the discussion of income effects from section 4.1, the mental accounting literature offers a more nuanced theory of income effects whereby individuals bracket different sources of income into mental accounts out of which they have different marginal propensities to consume (Thaler, 1985, 1999). Recent field studies have also found evidence of bracketing. In an in-store coupon field experiment involving an unanticipated coupon for a planned purchase, Heilman, Nakamoto, and Rao (2002) find that coupons cause more unplanned purchases of products that are related to the couponed item⁴⁸. Milkman and Beshears (2009) find that the incremental online consumer grocery purchases due to coupons are for non-typical items. Similarly, Hastings and Shapiro (2013) observe a much smaller cross-sectional correlation between household income and gasoline quality choice than the inter-temporal correlation between the gasoline price level

⁴⁷The empirical consumption literature has a long tradition of testing the extent to which consumer demand conforms with rationality by testing the integrability constraints associated with utility maximization (e.g., Lewbel, 2001; Hoderlein, 2011).

⁴⁸Lab evidence has also confirmed that consumers are much more likely to spend store gift card money on products associated with the brand of the card than unbranded gift card money (e.g. American Express), suggesting that store gift card money is not fungible with cash Reinholz, Bartels, and Parker (2015).

and gasoline quality choice. In related work, Hastings and Shapiro (2016) find that the income-elasticity of SNAP⁴⁹-eligible food demand is much higher with respect to SNAP benefits than with respect to cash. Each of these examples is consistent with consumers perceiving money budgeted for a product category differently from “cash.”

Hastings and Shapiro (2016) test the non-fungibility of income more formally using a demand model with income effects. Consider the bivariate utility over a commodity group, with J perfect substitutes products, and a $J + 1$ essential numeraire, with quadratic utility⁵⁰

$$U(\mathbf{x}) = \sum_{j=1}^J \psi_j x_j + \psi_{J+1,1} x_{J+1} - \frac{1}{2} \psi_{J+1,2} x_{J+1}^2$$

where $\psi_j = \exp(\bar{\psi}_j + \varepsilon_j)$. In the application, the goods consist of different quality grades of gasoline. In this model, incomes effects only arise through the allocation of the budget between the gasoline commodity group and the essential numeraire.

WLOG, if product k is the preferred good and, hence, $\frac{\psi_k}{p_k} = \min \left\{ \frac{p_j}{\psi_j} \right\}_{j=1}^J$, then the KKT conditions are

$$\psi_k - \psi_{J+1,1} p_k + \psi_{J+1,2} (y - x_k p_k) p_k \leq 0 \quad . \quad (79)$$

Estimation of this model follows from section 3.2. A simple test of fungibility consists of re-writing the KKT conditions with a different marginal utility on budget income and commodity expenditure

$$\left(\frac{\psi_k}{p_k} - \psi_{J+1,1} \right) + \psi_{J+1,y} y - \psi_{J+1,x} x_k p_k \leq 0 \quad . \quad (80)$$

and testing the hypothesis $H_0 : \psi_{J+1,y} = \psi_{J+1,x}$. The identification of this test relies on variation in both observed consumer income, y , and in prices, p .

⁴⁹SNAP refers to the Supplemental Nutrition Assistance Program, or “food stamps.”

⁵⁰Hastings and Shapiro (2013) treat quantities as exogenous and instead focus on the multinomial discrete choices problem between different goods, which are qualities of gasoline.

5.3.2 Social Preferences

As discussed in the survey by DellaVigna (2017), there is a large literature that has estimated social preferences in lab experiments. We focus herein specifically on the role of consumer's social preference and their responses to *cause marketing* campaigns involving charitable giving. A dominant theme of this literature has consisted of testing whether consumer response to charitable giving campaigns reflects genuine altruistic preferences versus alternative impure altruism and/or self-interest.

In a pioneering study, DellaVigna, List, and Malmendier (2012) conducted a door-to-door fundraising campaign to test the extent to which charitable giving is driven by a genuine preference to give (altruism or warm glow) versus a disutility from declining to give due, for instance, to social pressure. The field data are then used to estimate a structural model of individuals' utility from giving that separates altruism and social pressure.

Formally, total charitable giving, x consists of the sum of dollars donated to the charitable campaign either directly to the door-to-door solicitor, x_1 , or, if the donor is not home at the time of the visit, she can instead make a private donation, x_2 , by mail at an additional cost $(1 - \theta)x_2 \geq 0$ for postage, envelope etc. All remaining wealth is spent on an essential numeraire, x_3 , to capture all other private consumption. Propsective donors have a quasi-linear, bivariate utility over other consumption and charitable giving

$$U(x_1, x_2) = y - x_1 - x_2 + \tilde{U}(x_1 + \theta x_2) - s(x_1). \quad (81)$$

To ensure the sub-utility over giving, $\tilde{U}(x)$, (or “altruism” utility) has the usual monotonicity and concavity conditions, we assume $\tilde{U}(x) = \psi \log(\Gamma + x)$ where ψ is an altruism parameter and $\Gamma > 0$ influences the degree of concavity⁵¹. By allowing ψ to vary freely, the model captures the possibility of a donor who dislikes the charity. The third term in 81, $s(x)$, represents the social cost of declining to donate or giving a small donation to the solicitor. We assume $s(x) = \max(0, s(g - x))$

⁵¹Note that the marginal utility of giving $\frac{d\tilde{U}}{dx} = \frac{\psi}{\Gamma+x}$ so that high Γ implies a slow satiation on giving.

to capture the notion that the donor only incurs social pressure from donation amounts to the solicitor of less than g .

To identify the social preferences, DellaVigna, List, and Malmendier (2012) randomize subjects into several groups. In the first group, the solicitor shows up unannounced at the prospective donor's door. In this case, if the donor is home (with exogenous probability $h_0 \in (0, 1)$), she always prefers to give directly to the solicitor to avoid the additional cost $(1 - \theta)$ of donating by mail. The total amount given depends on the relative magnitudes of ψ and the social cost s . If the donor is not home, the only reason for her to donate via mail is due to altruism.

In the second group, the prospective donor is notified in advance of the solicitor's visit with a flyer left on the door. In this case, the donor can opt out by adjusting her probability of being home according to a cost $c(h - h_0) = \frac{(h - h_0)^2}{2\eta}$. The opt-out decision reflects the donor's trade-off between the utility of donating to the solicitor, subject to social pressure costs, and donating by mail, subject to mailing costs and the cost of leaving home. In a third group, subjects are given a costless option to "opt out" by checking a "do not disturb" box on the flyer, effectively setting $c(0) = 0$.

The authors estimate the model with a minimum distance estimator based on specific empirical moments from various experimental cells, although a maximum likelihood procedure might also have been used by including an additional random utility term into the model. While DellaVigna, List, and Malmendier (2012)'s estimates indicate that donations are driven by both social costs and altruism, the social cost estimates are surprisingly large. Almost half of the sample is found to prefer not to have a solicitation, either because they prefer not to donate or to donate a small amount. The results suggest a substantial welfare loss to donors from door-to-door solicitations. Moreover, the results indicate that the observed levels of charitable giving may not reflect altruism per se.

Kang, Park, Lee, Kim, and Allenby (2016) build on DellaVigna, List, and Malmendier (2012) by modeling the use of advertising creative to moderate the potential crowding-out effects of other

donations by others. They estimate a modified version of 81 for a prospective donor

$$U(x; G, \theta) = \theta_1 \ln(x + 1) + \theta_2 \ln(G) + \theta_3 \ln(y - x + 1) \quad (82)$$

where $G = x + x_{-i}$ measures total giving to the cause and x_{-i} represents the total stock of past donations from other donors. The authors can then test pure altruism, $\theta_1 = 0$ and $\theta_2 > 0$, versus a combination of altruism and warm-glow, $\theta_1 > 0$ and $\theta_2 > 0$ (Andreoni, 1989). The authors allow the relative role of warm glow to altruism, $\frac{\theta_1}{\theta_2}$, to vary with several marketing message variables. Since the preferences in 82 follow the Stone Geary functional form, demand estimation follows the approach in section 3.1.3 above.

The authors conduct a charitable giving experiment in which subjects were randomly assigned to different cells that varied the emotional appeals of the advertising message and also varied the reported amount of money donated by others. As in earlier work, the authors find that higher donations by others crowd out a prospective donor's contribution (Andreoni, 1989). The authors also find that recipient-focused advertising messages with higher arousal trigger the impure altruism appeal, which increases the level of donations.

Dubé, Luo, and Fang (2017) test an alternative self-signaling theory of crowding-out effects in charitable giving based on consumer's self-perception of altruistic preferences (Bodner and Prelec, 2002; Benabou and Tirole, 2006). Consumers make a binary purchase decision $x \in \{0, 1\}$ for a product with a price, p , and a pro-social characteristic $a \geq 0$ that measures the portion of the price that will be donated to a specific charity. Consumers obtain consumption utility from buying the product, $(\theta_0 + \theta_1 a + \theta_2 p)$ where θ_1 is the consumer's social preference or marginal utility for the donation. Consumers make the purchase in a private setting (e.g. online or on a mobile phone) with no peer influence (e.g., sales person or solicitor). In addition to the usual consumption utility, the consumer is uncertain about her own altruism and derives additional ego utility from the inference she makes about herself based on her purchase decision: $\theta_3 E(\theta_1 | a, p, x)$. θ_3 measures the consumer's ego utility⁵². The consumer chooses to buy if the combination of her consumption

⁵²In a social setting, this ego utility could instead reflect the value a consumer derives from conveying a "favorable

utility and ego utility exceed the ego utility derived from not purchasing:

$$(\theta_0 + \theta_1 a + \theta_2 p + \varepsilon_1) + \theta_3 E(\theta_1 | a, p, 1) > \varepsilon_0 + \theta_3 E(\theta_1 | a, p, 0) \quad (83)$$

where ε are choice-specific random utility shocks and θ_3 is the marginal ego utility associated with the consumer's self-belief about her own altruism, θ_1 . In this self-signaling model, the consumer's decision is driven not only by the maximization of consumption utility, but also by the equilibrium signal the consumer derives from her own action. Purchase and non-purchase have differential influences on the consumer's derived inference about her own ego utility, $E(\theta_1 | a, p, 0)$.⁵³

If $\tilde{\varepsilon} = \varepsilon_1 - \varepsilon_0 \sim N(0, \sigma^2)$, then consumer choice follows the standard random coefficients probit model of demand with purchase probability conditional on receiving the offer (a, p)

$$Pr(x = 1 | a, p) = \int \Phi(\theta_0 + \theta_1 a + \theta_2 p + \theta_3 [E(\theta_1 | a, p, 1) - E(\theta_1 | a, p, 0)]) dF(\theta) \quad (84)$$

where $F(\theta)$ represents the consumer's beliefs about her own preferences prior to receiving the ticket offer. Note that low prices can dampen the consumer's self-perception of being altruistic, $E(\theta_1 | a, p, 0)$, and reduce ego utility. If ego utility overwhelms consumption utility, consumer demand could exhibit backward-bending regions that would be inconsistent with the standard neo-classical framework.

Dubé, Luo, and Fang (2017) test the self-signaling theory through a cause marketing field experiment in partnership with a large telecom company and a movie theater. Subject received text messages with randomly-assigned actual discount offers for movie tickets. In addition, some subjects were informed that a randomized portion of the ticket price would be donated to a charity. In the absence of a donation, demand is decreasing in the net price. In the absence of a discount, demand is increasing in the donation amount. However, when the firm uses both a discount and a donation, the observed demand exhibits regions of non-monotonicity where the purchase rate

impression" (i.e. signal) to her peers based on her observed action.

⁵³As in Benabou and Tirole (2006), Dubé, Luo, and Fang (2017) also include $E(\theta_2 | a, p, x)$ in the ego utility to moderate the posterior belief by the consumer's self-perception of being sensitive to money.

declines at larger discount levels. These non-standard moments are used to fit the self-signaling model above in equation (84). The authors find that consumer response to the caus marketing campaign is driven more by ego utility, θ_3 , than by standard consumption utility.

6 Conclusions

Historically, the computational complexity of microeconomic models has limited their application to consumer-level transaction data. Most of the literature has focused on models of discrete brand choice, ignoring the more complicated aspects of demand for variety and purchase quantity decisions. Recent advances in computing power have mostly eliminated these computational challenges.

Many of the model specifications discussed herein require strong restrictions on preferences for analytic tractability, especially in the handling of corner solutions. These restrictions often rule out interesting and important aspects of consumer behavior such as income effects, product complementarity and indivisibility. We view the development of models to accommodate these richer behaviors as important directions for future research.

We also believe that the incorporation of ideas from behavioral economics and psychology into consumer models of demand will be a fruitful area for future research. Several recent papers have incorporated social preferences into traditional models of demand (e.g., DellaVigna, List, and Malmendier, 2012; Kang, Park, Lee, Kim, and Allenby, 2016; Dubé, Luo, and Fang, 2017). For a broader discussion of structural models of behavioral economics, see DellaVigna (2017).

Finally, the digital era has expanded the scope of consumer-level data available. These new databases introduce a new layer of complexity as the set of observable consumer features grows, sometimes into the thousands. Machine learning and regularization techniques offer potential opportunities for accommodating large quantities of potential variables into microeconomic models of demand. Devising approaches to conduct inference on structural models that utilize machine learning techniques will also likely offer an interesting opportunity for new research (e.g., Shiller

(2015) and Dubé and Misra (2017)).

This growing complexity due to indivisibilities, non-standard consumer behavior from the behavioral economics literature, and the size and scope of so-called “Big Data” raise some concerns about the continued practicality of the neoclassical framework for future research.

References

- AGUIAR, M., AND E. HURST (2007): “Life-Cycle Prices and Production,” *American Economic Review*, 97(5), 1533–1559.
- AINSLIE, A., AND P. E. ROSSI (1998): “Similarities in Choice Behavior Across Product Categories,” *Marketing Science*, 17(2), 91–106.
- ALLCOTT, H., R. DIAMOND, J. DUBE, J. HANDBURY, I. RAHKOVSKY, AND M. SCHNELL (2018): “Food Deserts and the Causes of Nutritional Inequality,” *Working Paper*.
- ALLENBY, G., M. J. GARRATT, AND P. E. ROSSI (2010): “A Model for Trade-Up and Change in Considered Brands,” *Marketing Science*, 29(1), 40–56.
- ALLENBY, G. M., AND P. E. ROSSI (1991): “Quality Perceptions and Asymmetric Switching between Brands,” *Marketing Science*, 10(3), 185–204.
- ALLENBY, G. M., T. S. SHIVELY, S. YANG, AND M. J. GARRATT (2004): “A Choice Model for Packaged Goods: Dealing with Discrete Quantities and Quantity Discounts,” *Marketing Science*, 23(1), 95–108.
- ANDERSON, S. P., AND A. DE PALMA (1992): “The Logit as a Model of Product Differentiation,” *Oxford Economic Papers*, 44, 51–67.
- ANDERSON, S. P., A. DE PALMA, AND J.-F. THISSE (1992): *Discrete Choice Theory of Product Differentiation*. The MIT Press.
- ANDREONI, J. (1989): “Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence,” *Journal of Political Economy*, 97, 1447–1458.
- ARORA, N. A., G. ALLENBY, AND J. L. GINTER (1998): “A Hierarchical Bayes Model of Primary and Secondary Demand,” *Marketing Science*, 17, 29–44.
- BECKER, G. S. (1965): “A Theory of the Allocation of Time,” *The Economic Journal*, 75(299), 493–517.
- BELL, D. R., J. CHIANG, AND V. PADMANABHAN (1999): “The Decomposition of Promotional Response: An Empirical Generalization,” *Marketing Science*, 18(4), 504–526.
- BENABOU, R., AND J. TIROLE (2006): “Incentives and Prosocial Behavior,” *American Economic Review*, 96, 1652–1678.
- BERRY, S., A. GANDHI, AND P. HAILE (2013): “Connected Substitutes and Invertibility of Demand,” *Econometrica*, 81(5), 2087–2111.

- BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): “Automobile Prices in Market Equilibrium,” *Econometrica*, 63(4), 841–890.
- BERRY, S. T. (1994): “Estimating Discrete-Choice Models of Product Differentiation,” *Rand Journal of Economics*, 25(2), 242–262.
- BESANKO, D., M. K. PERRY, AND R. H. SPADY (1990): “The Logit Model of Monopolistic Competition: Brand Diversity,” *Journal of Industrial Economics*, 38(4), 397–415.
- BHAT, C. R. (2005): “A Multiple Discrete-Continuous Extreme Value Model: Formulation and application to discretionary time-use decisions,” *Transportation Research Part B*, 39, 679–707.
- (2008): “The Multiple Discrete-Continuous Extreme Value (MDCEV) Model: Role of utility function parameters, identification considerations, and model extensions,” *Transportation Research Part B*, 42, 274–303.
- BHAT, C. R., M. CASTRO, AND A. R. PINJARI (2015): “Allowing for complementarity and rich substitution patterns in multiple discrete-continuous models,” *Transportation Research Part B*, 81, 59–77.
- BLATTBERG, R., T. BUESING, P. PEACOCK, AND S. SEN (1978): “Identifying the Deal Prone Segment,” *Journal of Marketing Research*, 15(3), 369–377.
- BLATTBERG, R. C., R. BRIESCH, AND E. J. FOX (1995): “How Promotions Work,” *Marketing Science*, 14(3), G122–G132.
- BLATTBERG, R. C., G. D. EPPEN, AND J. LIEBERMAN (1981): “A Theoretical and Empirical Evaluation of Price Deals for Consumer Nondurables,” *Journal of Marketing*, 45, 116–129.
- BLATTBERG, R. C., AND S. A. NESLIN (1989): “Sales Promotion: The Long and Short of It,” *Marketing Letters*, 1(1), 81–97.
- BLATTBERG, R. C., AND K. J. WISNIEWSKI (1989): “Price-Induced Patterns of Competition,” *Marketing Science*, 8, 291–309.
- BLUNDELL, R., P. PASHARDES, AND G. WEBER (1993): “What do we Learn About Consumer Demand Patterns from Micro Data?,” *American Economic Review*, 83(3), 570–597.
- BODNER, R., AND D. PRELEC (2002): *Collected Essays in Psychology and Economics*. Self-signaling and diagnostic utility in everyday decision making. Oxford University Press.
- BRONNENBERG, B. J., S. K. DHAR, AND J.-P. DUBÉ (2005): “Market Structure and the Geographic Distribution of Brand Shares in Consumer Package Goods Industries,” *manuscript*.

- BRONNENBERG, B. J., AND J.-P. DUBÉ (2017): “The Formation of Consumer Brand Preferences,” *Annual Review of Economics*, 9, 353–82.
- CARDELL, N. S. (1997): “Variance Components Structures for the Extreme-Value and Logistic Distributions with Applications to Models of Heterogeneity,” *Econometric Theory*, 13(2), 185–213.
- CHAMBERS, C. P., AND F. ECHENIQUE (2009): “Supermodularity and preferences,” *Journal of Economic Theory*, 144, 1004–1014.
- CHAMBERS, C. P., F. ECHENIQUE, AND E. SHMAYA (2010): “On behavioral complementarity and its implications,” *Journal of Economic Theory*, 145(6), 2332–2355.
- CHIANG, J. (1991): “A Simultaneous Approach to the Whether, What and How Much to Buy Questions,” *Marketing Science*, 10, 297–315.
- CHIANG, J., AND L.-F. LEE (1992): “Discrete/continuous models of consumer demand with binding nonnegativity constraints,” *Journal of Econometrics*, 54, 79–93.
- CHIB, S., P. SEETHARAMAN, AND A. STRIJNEV (2002): *Advances in Econometrics: Econometric Models in Marketing* chap. Analysis of Multi-Category Purchase Incidence Decisions Using IRI Market Basket Data, pp. 57–92. JAI Press.
- CHINTAGUNTA, P. (1993): “Investigating Purchase Incidence, Brand Choice and Purchase Quantity Decisions of Households,” *Marketing Science*, 12, 184–208.
- CHINTAGUNTA, P., J.-P. DUBÉ, AND K. Y. GOH (2005): “Beyond the Endogeneity Bias: The Effect of Unmeasured Brand Characteristics on Household-Level Brand Choice Models,” *MANAGEMENT SCIENCE*, 51, 832–849.
- CHINTAGUNTA, P., AND J.-P. DUBÉ (2005): “Estimating a Stockkeeping-Unit-Level Brand Choice Model That Combines Household Panel Data and Store Data,” *Journal of Marketing Research*, XLII(August), 368–379.
- CONLEY, T. G., C. B. HANSEN, R. E. MCCULLOCH, AND P. E. ROSSI (2008): “A semi-parametric Bayesian approach to the instrumental variable problem,” *Journal of Econometrics*, 144(1), 276–305.
- DEATON, A., AND J. MUELLBAUER (1980a): “An Almost Ideal Demand System,” *American Economic Review*, 70(3), 312–326.
- (1980b): *Economics and Consumer Behavior*. Cambridge University Press.
- DELLAVIGNA, S. (2017): *forthcoming Handbook of Behavioral Economics* chap. Structural Behavioral Economics. North-Holland.

- DELLAVIGNA, S., J. LIST, AND U. MALMENDIER (2012): “Testing for Altruism and Social Pressure in Charitable Giving,” *Quarterly Journal of Economics*, 127(1), 1–56.
- DOLAN, R. J. (1987): “Quantity Discounts: Managerial Issues and Research Opportunities,” *Marketing Science*, 6(1), 1–22.
- DOTSON, J. P., J. R. HOWELL, J. D. BRAZELL, T. OTTER, P. J. LENK, S. MACEACHERN, AND G. ALLENBY (2018): “A Probit Model with Structured Covariance for Similarity Effects and Source of Volume Calculations,” *Journal of Marketing Research*, 55, 35–47.
- DRAGANSKA, M., AND D. C. JAIN (2004): “A Likelihood Approach to Estimating Market Equilibrium Models,” *Management Science*, 50(5), 605–616.
- DU, R. Y., AND W. A. KAMAKURA (2008): “Where Did All That Money Go? Understanding How Consumers Allocate Their Consumption Budget,” *Journal of Marketing*, 72(November), 109–131.
- DUBÉ, J.-P. (2004): “Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks,” *Marketing Science*, 23(1), 66–81.
- DUBÉ, J.-P., G. HITSCH, AND P. ROSSI (2017): “Income and Wealth Effects on Private Label Demand: Evidence From the Great Recession,” *forthcoming Marketing Science*.
- DUBÉ, J.-P., X. LUO, AND Z. FANG (2017): “Self-Signaling and Pro-Social Behavior: a cause marketing mobile field experiment,” *Marketing Science*, 36(2), 161–186.
- DUBÉ, J.-P., AND S. MISRA (2017): “Scalable Price Targeting,” *Chicago Booth School of Business Working Paper*.
- DUBOIS, P., R. GRIFFITH, AND A. NEVO (2014): “Do Prices and Attributes Explain International Differences in Food Purchases?,” *American Economic Review*, 2014(3), 832–867.
- EINAV, L., E. LEIBTAG, AND A. NEVO (2010): “Recording Discrepancies in Nielsen Homescan Data: Are They Present and Do They Matter?,” *Quantitative Marketing and Economics*, 8(2), 207–239.
- ENGEL, E. (1857): “Die Productions-und Consumtionsver-haltnisse des Konigreichs Sachsen,” *Zeitschrift des Statistischen Bureaus des Koniglich Sachsischen Ministeriums des Innern*, 8, 1–54.
- ERDEM, T. (1998): “An Empirical Analysis of Umbrella Branding,” *Journal of Marketing Research*, 35(3), 339–351.
- ERDEM, T., S. IMAI, AND M. P. KEANE (2003): “Brand and Quantity Choice Dynamics Under Price Uncertainty,” *Quantitative Marketing and Economics*, 1, 5–64.

- ERDEM, T., M. P. KEANE, T. S. ÖNCÜ, AND J. STREBEL (2005): “Learning About Computers: An Analysis of Information Search and Technology Choice,” *Quantitative Marketing and Economics*, 3, 207–246.
- ERDEM, T., M. P. KEANE, AND B.-H. SUN (1999): “Missing price and coupon availability data in scanner panels: Correcting for the self-selection bias in choice model parameters,” *Journal of Econometrics*, 89, 177–196.
- GENTZKOW, M. (2007): “Valuing New Goods in a Model with Complementarity: Online Newspapers,” *American Economic Review*, 97(3), 713–744.
- GICHEVA, D., J. HASTINGS, AND S. B. VILLAS-BOAS (2010): “Investigating Income Effects in Scanner Data: Do Gasoline Prices Affect Grocery Purchases?,” *American Economic Review: Papers & Proceedings*, 100, 480–484.
- GOETTLER, R. L., AND K. CLAY (2011): “tariff Choice with Consumer learning and Switching Costs,” *Journal of Marketing Research*, XLVIII(August), 633–652.
- GOOLSBEE, A., AND A. PETRIN (2004): “The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV,” *Econometrica*, 72(2), 351–381.
- GUADAGNI, P. M., AND J. D. LITTLE (1983): “A Logit Model of Brand Choice Calibrated on Scanner Data,” *Marketing Science*, 2, 203–238.
- GUPTA, S. (1991): “Stochastic Models of Interpurchase Time with Time-Dependent Covariates,” *Journal of Marketing Research*, 28, 1–15.
- GUPTA, S., P. CHINTAGUNTA, A. KAUL, AND D. R. WITTINK (1996): “Do Household Scanner Data Provide Representative Inferences from Brand Choices: A Comparison with Store Data,” *Journal of Marketing Research*, 33(4), 383–398.
- HANEMANN, W. M. (1984): “Discrete/Continuous Models of Consumer Demand,” *Econometrica*, 52(3), 541–561.
- HARTMANN, W. R., AND H. S. NAIR (2010): “Retail Competition and the Dynamics of Demand for Tied Goods,” *Marketing Science*, 29(2), 366–386.
- HASTINGS, J., AND J. M. SHAPIRO (2013): “Fungibility and Consumer Choice: Evidence from Commodity Price Shocks,” *Quarterly Journal of Economics*, 128(4), 1449–1498.
- (2016): “How Are SNAP Benefits Spent? Evidence From a Retail Panel,” *Brown University Working Paper*.
- HAUSMAN, J. A. (1985): “The Econometrics of Nonlinear Budget Sets,” *Econometrica*, 53(6), 1255–1282.

- HECKMAN, J. J. (1978): “Dummy Endogenous Variables in a Simultaneous Equation System,” *Econometrica*, 46(4), 931–959.
- (1981): “the Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process and Some Monte Carlo Evidence,” in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski, and D. McFadden, chap. 4, pp. 179–195. MIT Press.
- HEILMAN, C. M., K. NAKAMOTO, AND A. G. RAO (2002): “Pleasant Surprises: Consumer Response to Unexpected In-Store Coupons,” *Journal of Marketing Research*, 39(2), 242–252.
- HENDEL, I. (1999): “Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns,” *Review of Economic Studies*, 66, 423–446.
- HENDEL, I., AND A. NEVO (2003): “The Post-Promotion Dip Puzzle: What do the Data Have to Say?,” *Quantitative Marketing and Economics*, 1, 409–424.
- (2006a): “Measuring the Implications of Sales and Consumer Inventory Behavior,” *Econometrica*, 74(6), 1637–1673.
- (2006b): “Sales and consumer inventory,” *RAND Journal of Economics*, 37(3), 543–561.
- HICKS, J., AND R. ALLEN (1934): “A Reconsideration of the Theory of Value. Part I,” *Economica*, 1(1), 52–76.
- HODERLEIN, S. (2011): “How many consumers are rational?,” *Journal of Econometrics*, 164, 294–309.
- HONKA, E. (2014): “Quantifying Search and Switching Costs in the U.S. Auto Insurance Industry,” *Rand Journal of Economics*, 45, 847–884.
- HOUTHAKER, H. (1953): “La Forme des Courbes,” *Cahiers du Seminaire d’Econometrie*, (2), 59–66.
- HOUTHAKKER, H. (1961): “The Present State of Consumption Theory,” *Econometrica*, 29(4), 704–740.
- HOWELL, J., AND G. ALLENBY (2017): “Choice Models with Fixed Costs,” *Working Paper*.
- HOWELL, J., S. LEE, AND G. ALLENBY (2016): “Price Promotions in Choice Models,” *Marketing Science*, 35(2), 319–334.
- JOO, J. (2018): “Quantity Surcharged Larger Package Sales as Rationally Inattentive Consumers’ Choice,” *University of Texas at Dallas Working Paper*.

- KANG, M. Y., B. PARK, S. LEE, J. KIM, AND G. ALLENBY (2016): “Economic Analysis of Charitable Donations,” *Journal of Marketing and Consumer Behaviour in Emerging Markets*, 2(4), 40–57.
- KAO, C., L. FEI LEE, AND M. M. PITT (2001): “Simulated Maximum Likelihood Estimation of the Linear Expenditure System with Binding Non-Negativity Constraints,” *Annals of Economics and Finance*, 2, 215–235.
- KIM, J., G. M. ALLENBY, AND P. E. ROSSI (2002): “Modeling Consumer Demand for Variety,” *Marketing Science*, 21(3), 229–250.
- (2007): “Product attributes and models of multiple discreteness,” *Journal of Econometrics*, 138, 208–230.
- LAMBRECHT, A., K. SEIM, AND B. SKIERA (2007): “Does Uncertainty Matter? Consumer Behavior Under Three-Part Tariffs,” *Marketing Science*, 26(5), 698–710.
- LEE, J., AND G. ALLENBY (2009): “A Direct Utility Model for Market Basket Data,” *OSU Working Paper*.
- LEE, L.-F., AND M. M. PITT (1986): “Microeconomic Demand Systems with Binding Non-negativity Constraints: The Dual Approach,” *Econometrica*, 5, 123–1242.
- LEE, R. S. (2013): “Vertical Integration and Exclusivity in Platform and Two-Sided Markets,” *American Economic Review*, 103(7), 2960–3000.
- LEE, S., AND G. ALLENBY (2014): “Modeling Indivisible Demand,” *Marketing Science*, 33(3), 364–381.
- LEE, S., J. KIM, AND G. M. ALLENBY (2013): “A Direct Utility Model for Asymmetric Complements,” *Marketing Science*, 32(3), 454–470.
- LEWBEL, A. A. (2001): “Demand systems with and without errors,” *American Economic Review*, 91(3), 611–618.
- LIU, Y., AND S. BALACHANDER (2014): “How long has it been since the last deal? Consumer promotion timing expectations and promotional response,” *Quantitative Marketing and Economics*, 12, 85–126.
- LUCE, R. D. (1977): “The Choice Axiom After Twenty Years,” *Journal of Mathematical Psychology*, 15, 215–233.
- MA, Y., K. L. AILAWADI, D. K. GAURI, AND D. GREWAL (2011): “An Empirical Investigation of the Impact of Gasoline Prices on Grocery Shopping Behavior,” *Journal of Marketing*, 75(2), 18–35.

- MANCHANDA, P., A. ANSARI, AND S. GUPTA (1999): "The "Shopping Basket": A Model for Multicategory Purchase Incidence Decisions," *Marketing Science*, 18, 95–114.
- MANCHANDA, P., P. E. ROSSI, AND P. CHINTAGUNTA (2004): "Response Modeling with Non-random Marketing-Mix Variables," *Journal of Marketing Research*, 41(4), 467–478.
- MANSKI, C. F., AND L. SHERMAN (1980): "An Empirical Analysis of Household Choice among Motor Vehicles," *Transportation Research*, 14(A), 349–366.
- MAS-COLLEL, A., M. D. WHINSTON, AND J. R. GREEN (1995): *Microeconomic Theory*. Oxford University Press.
- MATEJKA, F., AND A. MCKAY (2015): "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, 105(1), 272–298.
- MCCULLOCH, R., AND P. E. ROSSI (1994): "An exact likelihood analysis of the multinomial probit model," *Journal of Econometrics*, 64, 207–240.
- MCFADDEN, D. L. (1981): *Structural Analysis of Discrete Data and Econometric Applications* chap. 5, pp. 198–272. The MIT Press.
- MEHTA, N. (2007): "Investigating Consumers' Purchase Incidence and Brand Choice Decisions across Multiple Product Categories: A Theoretical and Empirical Analysis," *Marketing Science*, 26(2), 196–217.
- (2015): "A Flexible Yet Globally Regular Multigood Demand System," *Marketing Science*, 34(6), 843–863.
- MEHTA, N., X. J. CHEN, AND O. NARASIMHAN (2010): "Examining Demand Elasticities in Hanemann's Framework: A Theoretical and Empirical Analysis," *Marketing Science*, 29, 422–437.
- MEHTA, N., AND Y. MA (2012): "A Multicategory Model of Consumers' Purchase Incidence, Quantity, and Brand Choice Decisions: Methodological Issues and Implications on Promotional Decisions," *Journal of Marketing Research*, XLIX(August), 435–451.
- MELNIKOV, O. (2013): "Demand for Differentiated Durable Products: The Case of the U.S. Computer Printer Market," *Economic Inquiry*, 51(2), 1277–1298.
- MILKMAN, K. L., AND J. BESHEARS (2009): "Mental accounting and small windfalls: Evidence from an online grocer," *Journal of Economic Behavior & Organization*, 71, 384–394.
- MILLIMET, D. L., AND R. TCHERNIS (2008): "Estimating high-dimensional demand systems in the presence of many binding non-negativity constraints," *Journal of Econometrics*, 147, 384–395.

- MISRA, S. (2005): “Generalized Reverse Discrete Choice Models,” *Quantitative Marketing and Economics*, 3, 175–200.
- MUELLBAUER, J. (1974): “Household Composition, Engel Curves and Welfare Comparisons Between Households,” *European Economic Review*, 10, 103–122.
- NAIR, H. S., AND P. CHINTAGUNTA (2011): “Discrete-Choice Models of Consumer Demand in Marketing,” *Marketing Science*, 30(6), 977–996.
- NAIR, H. S., P. CHINTAGUNTA, AND J.-P. DUBÉ (2004): “Empirical Analysis of Indirect Network Effects in the Market for Personal Digital Assistants,” *Quantitative Marketing and Economics*, 2, 23–58.
- NARAYANAN, S., P. CHINTAGUNTA, AND E. J. MIRAVETE (2007): “The role of self selection, usage uncertainty and learning in the demand for local telephone service,” *Quantitative Marketing and Economics*, 5, 1–34.
- NEARY, J., AND K. ROBERTS (1980): “The Theory of Household Behavior Under Rationing,” *European Economic Review*, 13, 25–42.
- NEVO, A. (2001): “Measuring Market Power in the Ready-to-Eat Cereal Industry,” *Econometrica*, 69(2), 307–342.
- (2011): “Empirical Models of Consumer Behavior,” *Annual Review of Economics*, 3, 51–75.
- OGAKI, M. (1990): “The Indirect and Direct Substitution Effects,” *The American Economic Review*, 80(5), 1271–1275.
- OHASHI, H. (2003): “The Role of Network Effects in the US VCR Market, 1978-1986,” *Journal of Economics and Management Strategy*, 12(4), 447–494.
- PAKES, A. (2014): “Behavioral and Descriptive Forms of Choice Models,” *International Economic Review*, 55(3), 603–624.
- PAUWELS, K., S. SRINIVASAN, AND P. H. FRANSES (2007): “When Do Price Thresholds Matter in Retail Categories?,” *Marketing Science*, 26(1), 83–100.
- PESENDORFER, M. (2002): “Retail Sales: A Study of Pricing Behavior in Supermarkets,” *Journal of Business*, 75(1), 33–66.
- PETRIN, A., AND K. E. TRAIN (2010): “A Control Function Approach to Endogeneity in Consumer Choice Models,” *Journal of Marketing Research*, 47(1), 3–13.

- PHANEUF, D., AND V. SMITH (2005): *Handbook of Environmental Economics*chap. Recreation Demand Models, pp. 671–762. North-Holland.
- POLLAK, R. A., AND T. J. WALES (1992): *Demand System Specification and Estimation*. Oxford University Press.
- RANSOM, M. R. (1987): “A Comment on Consumer Demand Systems with Binding Non-Negativity Constraints,” *Journal of Econometrics*, 34, 355–359.
- REINHOLTZ, N., D. BARTELS, AND J. R. PARKER (2015): “On the Mental Accounting of Restricted-Use Funds: How Gift Cards Change What People Purchase,” *Journal of Consumer Research*, 42, 596–614.
- REISS, P. C., AND M. W. WHITE (2001): “Household Electricity Demand Revisited,” *NBER Working Paper 8687*.
- SAMUELSON, P. A. (1974): “Complementarity: An Essay on the 40th Anniversary of the Hicks-Allen Revolution in Demand Theory,” *Journal of Economic Literature*, 12(4), 1255–89.
- SCOTT A. NESLIN, C. H., AND J. QUELCH (1985): “Consumer Promotions and the Acceleration of Product Purchases,” *Marketing Science*, 4(2), 147–165.
- SEETHARAMAN, P. B., S. CHIB, A. AINSLIE, P. BOATRIGHT, T. Y. CHAN, S. GUPTA, N. MEHTA, V. R. RAO, AND A. STRIJNEV (2005): “Models of Multi-Category Choice Behavior,” *Marketing Letters*, 16(3), 239–254.
- SEILER, S. (2013): “The Impact of Search Costs on Consumer Behavior: A Dynamic Approach,” *Quantitative Marketing and Economics*.
- SHILLER, B. R. (2015): “First-Degree Price Discrimination Using Big Data,” *Working Paper*.
- SONG, I., AND P. CHINTAGUNTA (2007): “A discrete-continuous model for multicategory purchase behavior of households,” *Journal of Marketing Research*, 44(November), 595–612.
- STONE, R. (1954): “Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand,” *The Economic Journal*, 255(64), 511–527.
- SUN, B. (2005): “Promotion Effect on Endogenous Consumption,” *Marketing Science*, 24(3), 430–443.
- SUN, B., S. A. NESLIN, AND K. SRINIVASAN (2003): “Measuring the Impact of Promotions on Brand Switching When Consumers Are Forward Looking,” *Journal of Marketing Research*, 40(4), 389–405.

- TAMER, E. (2003): “Incomplete Simultaneous Discrete Response Model with Multiple Equilibria,” *Review of Economic Studies*, 70, 147–165.
- (2010): “Partial Identification in Econometrics,” *Annual Review of Economics*, 2(1), 167–195.
- THALER, R. (1985): “Mental Accounting and Consumer Choice,” *Marketing Science*, 4(3), 199–214.
- THALER, R. H. (1999): “Anomalies: Saving, Fungibility, and Mental Accounts,” *Journal of Economic Perspectives*, 4(1), 193–205.
- THOMASSEN, O., S. SEILER, H. SMITH, AND P. SCHIRALDI (2017): “Multi-Category Competition and Market Power: A Model of Supermarket Pricing,” *forthcoming American Economic Review*.
- TRAIN, K. E., D. L. MCFADDEN, AND M. BEN-AKIVA (1987): “The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices,” *Rand Journal of Economics*, 18(1), 109–123.
- VAN SOEST, A., A. KAPTEYN, AND P. KOOREMAN (1993): “Coherency and regularity of demand systems with equality and inequality constraints,” *Journal of Econometrics*, 57, 161–188.
- VAN SOEST, A., AND P. KOOREMAN (1990): “Coherency of the Indirect Translog Demand System with Binding Nonnegativity Constraints,” *Journal of Econometrics*, 44, 391–400.
- VARIAN, H. R. (1989): *Price Discrimination* chap. 10, pp. 597–654. Elsevier Science Publishers BV.
- VILLAS-BOAS, J. M., AND R. S. WINER (1999): “Endogeneity in Brand Choice Models,” *Management Science*, 45(10), 1324–1338.
- VILLAS-BOAS, J. M., AND Y. ZHAO (2005): “Retailer, Manufacturers, and Individual Consumers: Modeling the Supply Side in the Ketchup Marketplace,” *Journal of Marketing Research*, XLII(February), 83–95.
- WALES, T., AND A. WOODLAND (1983): “Estimation of consumer demand systems with binding non-negativity constraints,” *Journal of Econometrics*, 21, 263–285.
- WALSH, J. W. (1995): “Flexibility in Consumer Purchasing for Uncertain Future Tastes,” *Marketing Science*, 14(2), 148–165.
- WOOLDRIDGE, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

YANG, S., Y. CHEN, AND G. M. ALLENBY (2003): "Bayesian analysis of simultaneous demand and supply," *Quantitative Marketing and Economics*, 1, 251–275.

YAO, S., C. F. MELA, J. CHIANG, AND Y. CHEN (2012): "Determining Consumers' Discount Rates with Field Studies," *Journal of Marketing Research*, 49, 822–841.