Non-Parametric Identification and Estimation of Demand and Preferences using Scanner Data

by

Christopher R. Dobronyi

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Department of Economics
University of Toronto

Abstract

Non-Parametric Identification and Estimation of Demand
and Preferences using Scanner Data

Christopher R. Dobronyi
Doctor of Philosophy
Department of Economics
University of Toronto
2021

This thesis presents three essays on the partial non-parametric identification and estimation of demand and preferences using scanner data. In the first essay, I consider a critical real-world problem: the formal identification and estimation of food stamp fraud in the United States. It is shown that consumption choices can be used to bound fraud. The estimation procedure in this essay uses a standard assumption. In particular, it assumes that there exists a conditional quantile of consumption that coincides with an individual demand function. This assumption is strong, but useful for the structural estimation problem at hand. In the second and third essays, I propose flexible methods for estimating demand and preferences using scanner data that do not require this strong assumption. In the first proposed method, heterogeneity across consumers is introduced by assuming that the marginal rate of substitution is a random field. In such an environment, the theory of generalized functions can be used to test the integrability of expected demand at a parametric rate. If variation in preferences is small, then preferences can be recovered by approximating the relationship between preferences and demand with a first-order expansion and applying an analogue of the delta-method. In the second proposed method, individual-level heterogeneity is characterized by a distribution $\pi \in \Pi$, and the heterogeneity across consumers is characterized by a Dirichlet process F over the distributions in Π. Two frameworks for estimation are considered: a Bayesian framework in which F is known, and a hyperparametric (or empirical Bayesian) framework in which F is a member of a parametric family. Both methods are illustrated by applications to alcohol consumption.

Acknowledgements

Chapter 1: I would like to thank my supervisor, Christian Gouriéroux, for all of his valuable feedback. I would also like to thank Rahul Deb, Yoram Halevy, and Ismael Mourifié for many helpful discussions.

Chapter 2: This chapter was written jointly with Christian Gouriéroux. I would like to thank Rahul Deb, Florian Gunsilius, Yoram Halevy, Ismael Mourifié, Whitney Newey, Yuanyuan Wan, and participants at the 31^{st} EC^2 Conference on High Dimenional Modeling in Time Series (2020), Econometric Society World Congress (2020), and a seminar at the University of Toronto for many helpful comments.

Chapter 3: This chapter was also written jointly with Christian Gouriéroux. I would like to thank Martin Burda, JoonHwan Cho, Rahul Deb, Jiaying Gu, Joann Jasiak, Kyoo il Kim, Yao Luo, Etienne Masson-Makdissi, Ismael Mourifié, Jesse Shapiro, Eduardo Souza-Rodrigues, as well as participants at the 14^{th} International Conference on Computational and Financial Econometrics (2020) and seminars at the University of Ottawa, University of Toronto, and University of Waterloo for many helpful comments.

I am grateful to Marine Carrasco for acting as an external member of my thesis committee, and providing me with extremely useful feedback.

All chapters were supported by Ontario Graduate Scholarships in 2018 and 2020. This thesis was also enabled by support from Sci-Net and Compute Canada.

The empirical results in this thesis are derived using marketing databases provided by the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. These results do not reflect the views of Nielsen.

To my partner and my family.

Contents

List of Tables

List of Figures

Chapter 1

Food Stamp Fraud

In the United States, poor households are given food stamps. Food stamps can only be used to buy food. Some households illegally exchange food stamps for cash. This type of exchange is called food stamp fraud (or food stamp trafficking). The Food and Nutrition Service (FNS) is responsible for detecting and analyzing food stamp fraud, but their methodology is limited, unable to detect certain types of fraud, and unable to assess the effect of a change in policy on welfare or fraud. In this chapter, I use scanner data to identify and non-parametrically estimate a model of food stamp fraud. This chapter is the first analysis in economics that attempts to solve this critical real-world problem.

The estimation procedure in this chapter is simple and standard. It assumes that there is a one-to-one relationship between demand (in the absence of food stamps) and preferences, and that the conditional quantiles of consumption coincide with individual demand functions. This assumption drastically reduces the complexity of the estimation problem. I use the analysis in this chapter to motivate the more flexible methods described in the following chapters.

1.1 Introduction

The Supplemental Nutrition Assistance Program (SNAP) is a federal aid program in the United States that provides low-income households with benefits that can be used to buy food. Roughly 1.5% of food stamps—equivalently, $1.1 billion in food stamps—are thought to be illegally exchanged with retailers for cash each year (Willey et al., 2017). This type of illegal exchange is called *food stamp fraud* or *food stamp trafficking*. It is thought to have increased by nearly 500% since 2002 (Willey et al., 2017).

Food stamp fraud is a concern because it redirects aid from low-income households to retailers. It is also a signal of inefficiency, motivating other forms of aid: If food stamp fraud exists, we can use *in-cash transfers* to increase welfare and food expenditure,[1] and decrease the cost of SNAP, without changing

[1]The objective of SNAP is to ensure that households are protected from malnutrition as a result of insufficient income, as described in the Food and Nutrition Act of 2008. Benefits are used, instead of, for instance, in-cash transfers, to ensure

incentives or increasing fraud.

The Food and Nutrition Service (FNS) is responsible for detecting and analyzing food stamp fraud. Fraud is detected using undercover investigations, social media, tips and referrals, and transaction data (Aussenberg, 2018), then analyzed using the data from these sources (Willey et al., 2017). This analysis is expensive. It also has limitations because it uses data associated with suspicious behaviour (see Section 2.2 in Willey et al., 2017), and it is unable to assess the effect of a change in policy on welfare or fraud.

In this chapter, I non-parametrically identify and estimate a model of food stamp fraud. My methodology is intended to replace, or serve as a complement to, the existing methodology used by the FNS. The analysis in this chapter uses economic theory to discern the necessary implications of fraud on consumption, then looks for these implications in the Panel Survey of Income Dynamics (PSID), and the Nielsen Homescan Consumer Panel (NHCP).

In the PSID, approximately 2 percent of households with benefits report being disqualified for breaking the rules, approximately 2 to 10 percent of households with benefits report expenditures that are consistent with fraud, and the average amount of benefits exchanged ranges from approximately \$325 to \$1,800, depending on the year. I estimate just over \$1 billion dollars of food stamp fraud in 2017, which is consistent with Willey et al. (2017). I find that expenditures that are consistent with fraud are more likely to be reported by poorer households in the South with more members and younger female heads.

I show that there is evidence of fraud in the NHCP for households with the above characteristics. I use this data to non-parametrically estimate the structural objects of interest in the model of food stamp fraud, including (i) demand for goods (in the presence of fraud), (ii) bounds on food stamp amounts (which are unobserved in the NHCP), (iii) the expected cost of an illegal exhange to the household, and (iv) bounds on demand for fraud. These objects can be used to rule out fraud for certain households or bound the effect of a change in policy on fraud.

The remainder of this chapter is organized as follows. In the next section, I introduce the model of fraud. Sections 1.3 and 1.4 discuss non-parametric identification and estimation, respectively. In Section 1.5, I provide the applications to the PSID and NHCP. Section 1.6 concludes.

1.2 A Model of Benefit Fraud

In this section, I describe a simple model of benefit fraud. In this model, households[2] get benefits if they are eligible, and commit fraud if they have an incentive to do so. In practice, some eligible households might not get benefits (maybe because they never apply for benefits), and some households might not commit fraud even if they have an incentive to do so—the model in this section is intended to serve as a benchmark.

that aid is spent on food. If households exhaust their budgets and treat benefits like cash, then, in the absence of benefit fraud, benefits have the ability to increase food expenditure by *more* than in-cash transfers, which can be desirable for paternalistic reasons, its effect on health, its effect on healthcare costs, or its effect on the welfare of dependents.

[2] Agents make up a *household* if they live together and prepare food together (see Appendix 1.A.1).

1.2.1 Benefits

The objective of the FNS is to ensure that households are protected from malnutrition due to insufficient income, as described in the Food and Nutrition Act of 2008. Eligible households are given Electronic Benefit Transfer (EBT) cards. EBT cards are loaded with benefits each month, and can be used like debit cards to buy pre-approved goods.

Assumption 1.1.

(i) There are two aggregate goods: food and non-food.[3]

(ii) Income can be used to buy food and non-food.

(iii) Benefits can only be used to buy food.

Let $z \in \mathbb{R}^2_{++}$ denote a pair (y, p), where y denotes household income, and p denotes the price of food, both normalized by the price of non-food. The household gets a non-negative allotment of benefits. The amount that it gets depends on z. Let $b(z)$ denote this amount, also normalized by the price of non-food.

Assumption 1.2. The household receives a non-negative allotment of benefits $b(z)$.

Benefits $b(z)$ only depend on *normalized* income and prices. This restriction holds if, before any normalizations, the policy $b(\cdot)$ is invariant to homothetic changes in income and prices—in other words, if, before any normalizations, scaling income and prices by a common factor yields a proportional change in benefits. This restriction is natural for a policy intended to protect households from insufficient income: Without such a restriction, a change in the unit of money would lead to a change in purchasing power.

Assumption 1.3.

(i) The policy $b(\cdot)$ is non-increasing in y, given p, on $\mathbb{R}_{++}$.

(ii) Rich households do not receive benefits.

All else equal, poorer households need more aid. Assumption 1.3 says that richer households get fewer benefits than poorer households when faced with identical prices, and that rich households are not allocated any benefits. Assumption 1.3 implies that the policy $b(\cdot)$ has the form:

$$b(z) = \begin{cases} a(z), & \text{if } y \leq c(p), \\ 0, & \text{otherwise,} \end{cases} \tag{1.2.1}$$

for real-valued functions, $a(\cdot)$ and $c(\cdot)$, where $a(\cdot)$ is non-increasing in income y, given p. The form of the policy $b(\cdot)$ in (1.2.1) is consistent with the Food and Nutrition Act of 2008. In practice, a household

[3]The names of these aggregate goods are for exposition—the results in this paper can be applied to any in-kind program. See Appendix 1.A.8 for a comprehensive description of goods classified as food.

is eligible to receive benefits if its income y is below a threshold $c(p)$ that is proportional to the poverty line, adjusted for inflation, and, if it satisfies this criterion, the amount $a(z)$ that it gets is calculated by subtracting 70 percent of its income from a maximum amount, adjusted for inflation, encompassing the cost of a "thrifty" food plan for a household of its size (see Appendix 1.A for a detailed description including maximum amounts). Likely, only 70 percent of its income is subtracted to, in part, ensure that the policy $b(\cdot)$ does not distort the ranking of eligible households of the same size with respect to purchasing power in the population (so that a decrease in income will never make an eligible household better off). This type of restriction is often introduced to ensure that a policy is fair, and can be found in, for instance, tax and employment insurance policies. In the current setting, this type of restriction has an important implication: If the policy $b(\cdot)$ does not distort the ranking of any two households of the same size, then "total income"—in particular, $\psi(z) = y + b(z)$—is strictly increasing in income y, given p. To be clear, the amount $a(\cdot)$ and threshold $c(\cdot)$ implicitly depend on household size, which is omitted for exposition.

Assumption 1.4.

(i) The amount $a(\cdot)$ is continuously-differentiable on $\mathbb{R}^2_{++}$.

(ii) The threshold $c(\cdot)$ is continuous on $\mathbb{R}_{++}$.

(iii) Total income $\psi(z) = y + b(z)$ is strictly increasing in y, given p, on $\mathbb{R}_{++}$.

Assumption 1.4(i) implies that the policy $b(\cdot)$ is smooth wherever it is positive, ruling out, for instance, piecewise-linear policies with two or more discontinuities. Assumption 1.4(ii) rules out pathological cases with thresholds $c(\cdot)$ that "jump." Assumption 1.4(iii) impedes the policy $b(\cdot)$ from decreasing in income y too quickly. Under Assumption 1.4(i), Assumption 1.4(iii) holds if, and only if:

$$-1 < \frac{\partial b(z)}{\partial y}, \tag{1.2.2}$$

at every $z \in \mathbb{R}^2_{++}$ such that $y \neq c(p)$. While Assumption 1.4 is both fair and natural for any benefit policy, only a few of the results that follow need this assumption. It will be made clear when we need it.

Example 1.1. To illustrate Assumptions 1.3 and 1.4, let us consider a parametric piecewise-linear policy $b(\cdot)$. Formally, let us consider the policy $b(\cdot)$ defined by:

$$b(z) = \begin{cases} \gamma_1 - \gamma_2 y + \gamma_3 p, & \text{if } y \leq \frac{\gamma_1 + \gamma_3 p}{\gamma_2}, \\ 0, & \text{otherwise}, \end{cases} \tag{1.2.3}$$

for every $z \in \mathbb{R}^2_{++}$, in which $\gamma \in \mathbb{R}^3_{+}$ denotes a vector of non-negative numbers with $\gamma_2 > 0$. Under this specification, the household gets benefits if, and only if, it cannot afford the pair consisting of γ_3 units of food and γ_1 units of non-food, after scaling income by γ_2. If it satisfies this criterion, it gets the exact amount of benefits needed to make the bundle (γ_3, γ_1) affordable, after scaling income by γ_2.

Now, notice, we can also write:

$$b(z) = \left[\gamma_3 p - \gamma_2 \left(y - \frac{\gamma_1}{\gamma_2}\right)\right]^+, \tag{1.2.4}$$

for every $z \in \mathbb{R}^2_{++}$. This reformulation yields an alternative interpretation. In particular, it is equivalent to say that, under this specification, the household gets benefits if, and only if, it cannot afford γ_3 units of food, after drifting income by γ_1/γ_2, and scaling income by γ_2. In both interpretations, we are scaling and/or drifting income, as done in practice. This parametric policy satisfies Assumption 1.3. It also satisfies Assumption 1.4 if, and only if, $\gamma_2 < 1$, since, for every $z \in \mathbb{R}^2_{++}$ such that $y < \frac{\gamma_1 + \gamma_3 p}{\gamma_2}$, we have:

$$\frac{\partial b(z)}{\partial y} = \frac{\partial}{\partial y}\left[\gamma_1 - \gamma_2 y + \gamma_3 p\right] = -\gamma_2. \tag{1.2.5}$$

While theoretical, this parametric policy is a real option for protecting the household from malnutrition due to insufficient income—it is, in fact, similar to the description of the policy in the Food and Nutrition Act of 2008 when $\gamma_1 = 0$ and $\gamma_2 = 0.7$. While restrictive, parametric forms can be used to find an optimal policy (within a family), or examine the effect of a change in the policy by means of a shock on the parameters. For simplicity, in subsequent examples, I will make use of the following "special case":

$$b(z) = \begin{cases} \gamma_1 - \gamma_2 y, & \text{if } y \leq \frac{\gamma_1}{\gamma_2}, \\ 0, & \text{otherwise,} \end{cases} \tag{1.2.6}$$

for every $z \in \mathbb{R}^2_{++}$, where $\gamma_2 > 0$. Under this specification, the household gets benefits if, and only if, its income y is smaller than γ_1/γ_2. If it satisfies this criterion, it gets the maximum amount γ_1 that a household can receive—the amount associated with an income y of zero—less income y scaled by γ_2. Intuitively, in this special case, the prices used in the calculation of the household's allocation have been "fixed" by the FNS. In practice, the household must immediately report all changes in income y, so that its allotment can be adjusted, but maximum amounts are only updated once a year. This policy often assigns more or less benefits than necessary. △

Assumption 1.4 has an implication that is worth discussing here:

Proposition 1.1. Under Assumptions 1.1 to 1.4, the policy $b(\cdot)$ is continuous in y on $\mathbb{R}^2_{++}$.

Proof. See Appendix 1.B.1. □

Proposition 1.1 says that the policy $b(\cdot)$ will not jump from a change in income y. While, in practice, jumps may exist, they should be avoided since they are unfair (as they distort the ranking of households with respect to purchasing power) and can create undesirable incentives for both workers and employers.

Example 1.2. To illustrate these undesirable incentives, let us consider yet another parametric piecewise-linear policy $b(\cdot)$. Specifically, let us consider the policy $b(\cdot)$ defined by:

$$b(z) = \begin{cases} \gamma_1 - \gamma_2 y, & \text{if } y \leq \gamma_3, \\ 0, & \text{otherwise,} \end{cases} \tag{1.2.7}$$

for every $z \in \mathbb{R}^2_{++}$, in which $\gamma \in \mathbb{R}^3_+$ denotes a vector of non-negative numbers with $\gamma_2 > 0$. This policy generalizes the policy with fixed prices in (1.2.6). It is more general because the threshold $c(p) = \gamma_3$ is independent of the amount $a(z) = \gamma_1 - \gamma_2 y$. This policy satisfies Assumption 1.3, but it violates Assumption 1.4 unless $\gamma_2 < 1$ and $\gamma_3 = \gamma_1/\gamma_2$, as in (1.2.6). It should also be clear that the policy b is continuous if these restrictions hold. To understand the claim about incentives, suppose $\gamma = (50, 0, 100)'$. This choice of the parameter γ implies that the household gets \$50 in benefits if, and only if, its income y is smaller than \$100. If the household makes \$20 per hour, then it will get \$50 in benefits if, and only if, it works 25 hours or less. If it works 24 hours, its total income ψ will be $24 \times \$20 + \$100 = \$580$; if it works 25 hours, its total income ψ will be $25 \times \$20 = \$500$. This discrepancy could lead to fewer hours worked. It could also lead to the employer decreasing its wage to get the household to work more for less. These incentives follow from a distortion in the ranking of total income. △

Example 1.3. The parametric policies in Examples 1.1 and 1.2 are very restrictive because they assume that the amount $a(\cdot)$ is linear in z. There is, however, no reason for us to impose linearity. Let us consider one last parametric policy $b(\cdot)$ defined by:

$$b(z) = \begin{cases} \gamma_1 - \gamma_2 y^2, & \text{if } y \leq \gamma_3, \\ 0, & \text{otherwise,} \end{cases} \tag{1.2.8}$$

for every $z \in \mathbb{R}^3_{++}$, in which $\gamma \in \mathbb{R}^2_+$ denotes a vector of non-negative numbers. This policy is similar to the policy in Example 1.2, but the amount $a(\cdot)$ is non-linear in income y. This policy satisfies both Assumptions 1.3 and 1.4 if, and only if, $\gamma_3 = \sqrt{\gamma_1/\gamma_2}$ and $0 < \gamma_2 < \sqrt{\gamma_2/4\gamma_1}$. To obtain this conclusion, notice that, this policy is continuous if, and only if, $\gamma_3 = \sqrt{\gamma_1/\gamma_2}$. Moreover, for every $z \in \mathbb{R}^2_{++}$ such that $y < \gamma_3$, the derivative of the policy $b(\cdot)$ equals:

$$\frac{\partial b(z)}{\partial y} = \frac{\partial}{\partial y}\left[\gamma_1 - \gamma_2 y^2\right] = -2\gamma_2 y. \tag{1.2.9}$$

This derivative is strictly larger than -1 at z if, and only if, $\gamma_2 < 1/2y$. Because this derivative is strictly decreasing in income y, the restriction in (1.2.2) holds if, and only if, $\gamma_2 < \sqrt{\gamma_2/4\gamma_1}$. The non-linearity of $a(\cdot)$ can be used to redistribute total income $\psi(\cdot)$ in the population. △

Remark 1.1. In practice, there are actually two types of thresholds that determine whether a household

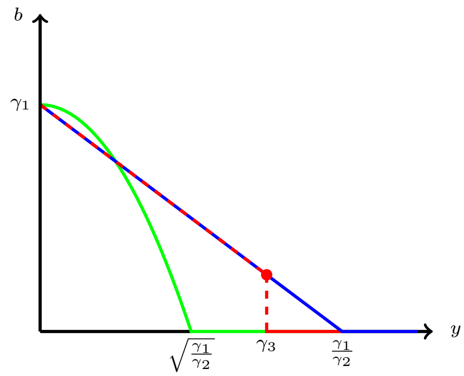

Figure 1.1. **Examples of Benefit Policies.** The blue line illustrates the policy with fixed prices in (1.2.6) from Example 1.1; the red line illustrates the policy in Example 1.2; the green line illustrates the policy in Example 1.3. In this figure, I have used: $\gamma_1 > \gamma_2$.

is eligible: the first type concerns the household's income, as described above; the second type concerns the household's "resources" (e.g., cash, savings, and personal vehicles). I neglect the second type for simplicity. See Appendix 1.A.2 for details.

1.2.2 Benefit Fraud

Let $\bar{R} = \mathbb{R}^2_+$ denote the consumption set—that is, the set of all possible "bundles." In the absence of benefit fraud, the household can afford a bundle $x \in \bar{R}$ if, and only if:

$$px_1 + x_2 \leq y + b(z) \quad \text{and} \quad x_2 \leq y, \tag{1.2.10}$$

where x_1 is a quantity of food, and x_2 is a quantity of non-food. The first inequality says that it can afford x only if the entire bundle x costs less than total income $\psi(z)$; the second inequality says that it can afford x only if the quantity x_2 costs less than income y. The quantity x_2 is constrained by income y because benefits cannot be used to buy non-food. The price of non-food in (1.2.10) is 1 since the price of food p, income y, and benefits $b(z)$ are relative to the price of non-food, as described in Section 1.2.1.

The household is not required to spend all of its income y or benefits $b(z)$. Formally, nothing is impeding the household from choosing a bundle x at which both of the inequalities in (1.2.10) are strict. That being said, under some weak conditions, the household will be constrained by at least one of these inequalities. If the household is constrained by the second inequality, it might have an incentive to use *benefit fraud* to afford a larger quantity of non-food.

Assumption 1.5.

(i) The household can exchange its benefits (or a portion of its benefits) for cash.

(ii) If the household exchanges f units of benefits, it gets πf in cash, for $\pi \in (0, 1)$.

Assumption 1.5(i) says that the household can exchange its benefits for cash. This type of exchange is a form of benefit fraud. This type of fraud often requires collusion—specifically, the household often requires an agent that is willing to provide cash in exchange for its benefits. This agent can be a second household (known in real life or met online using social media), but it is often a retailer. Assumption 1.5(ii) implies that, if the household exchanges $f \in [0, b(z)]$ units of benefits with a retailer, then it gets πf in (normalized) cash, and the retailer gets $(1-\pi)f$ in (normalized) profits. The discount factor π will be assumed to be exogenous to the household (see Assumption 1.7(i) in Section 1.2.5), determined by either the other agent or, quite possibly, an underlying market.

In practice, there are many ways for the household to commit benefit fraud: The household can (i) exchange benefits for cash, (ii) exchange benefits for ineligible goods, (iii) use benefits to buy food in a container with a return deposit with the intention of discarding its contents only to obtain the deposit (known as "water dumping" when the product is packaged water), (iv) sell its EBT card to a second agent, (v) purchase and then resell food, or (vi) lend a retailer its EBT card, so she can use it to buy food for her store (known as "indirect benefit fraud"). In recent years, the FNS has made the fourth way relatively infeasible by limiting the number of times that a household can replace its EBT card. Furthermore, the fifth way does not require any collusion. While Assumption 1.5(i) only describes the first way, each way has the same effect on the household's ability to buy goods. I focus on the first way only for exposition. All of these forms of fraud are illegal (see Appendix 1.A.10 for a formal definition).

Example 1.4. Suppose that the household gives a retailer \$100 in benefits, and, in return, the retailer gives the household \$50 in cash. This exchange can be completed by making a "fake" sale—for example, by charging the household for \$100 worth of food, and providing them with \$50 in cash, instead of food. In this example, we obtain:

$$f = 100 \quad \text{and} \quad \pi = \frac{50}{100} = 0.5. \tag{1.2.11}$$

△

Example 1.5. Suppose that the household uses \$10 in benefits to buy 24 water bottles and, then, dumps and returns each bottle for 10 cents. This example yields:

$$f = 10 \quad \text{and} \quad \pi = \frac{24 \times 0.10}{10} = 0.24. \tag{1.2.12}$$

In general, water dumping yields a small discount factor, making it a more extreme form of benefit fraud, only performed by desperate households. △

Example 1.6. Suppose that the household uses \$50 in benefits to buy 10 bags of roasted almonds and resells these bags for \$4 each over eBay. This example produces the following values:

$$f = 50 \quad \text{and} \quad \pi = \frac{10 \times 4}{50} = 0.8. \tag{1.2.13}$$

As previously mentioned, this way to commit benefit fraud does not require collusion since the second agent does not know that the almonds were bought with benefits. △

1.2.3 Budget Constraints

Under Assumption 1.5, if the household exchanges $f \in [0, b(z)]$ units of benefits, its income will increase to $y + \pi f$, and its benefits will decrease to $b(z) - f$. Hence, under the option of benefit fraud, it can afford $x \in \bar{R}$ if there exists $f \in [0, b(z)]$ such that:

$$px_1 + x_2 \leq \big[y + \pi f\big] + \big[b(z) - f\big] \;\text{ and }\; x_2 \leq y + \pi f. \tag{1.2.14}$$

These inequalities follow from replacing income with $y + \pi f$ and benefits with $b(z) - f$ in (1.2.10). Increasing f tightens the first inequality, and relaxes the second, making it possible for the household to use benefit fraud to afford a larger quantity of non-food.

Example 1.7. Suppose that the household has \$50 in income y and \$100 in benefits $b(z)$. Here, in the absence of benefit fraud, the household can afford a bundle $x \in \bar{R}$ if, and only if:

$$px_1 + x_2 \leq 150 \;\text{ and }\; x_2 \leq 50. \tag{1.2.15}$$

Further suppose that the discount factor π is 0.5, as in Example 1.4. If the household exchanges all \$100 of its benefits, its income will increase to $\$50 + 0.5 \times \$100 = \$100$, and its benefits will decrease to $\$100 - \$100 = \$0$. Thus, after exchanging all of its benefits, it can afford a bundle $x \in \bar{R}$ if, and only if:

$$px_1 + x_2 \leq 100 \;\text{ and }\; x_2 \leq 100. \tag{1.2.16}$$

△

1.2.4 Preferences

The household has preferences over the bundles in the consumption set $\bar{R}$. Its preferences are summarized by a utility function $u(\cdot)$ on $\bar{R}$. For each $v \in \mathbb{R}$, let $G(v) = \{x \in \bar{R} : u(x) = v\}$ denote the subset of bundles that attain a utility level of v. Let R denote the interior of $\bar{R}$.

Assumption 1.6.

(i) Utility $u(\cdot)$ is twice-continuously-differentiable on R.

(ii) Utility $u(\cdot)$ is strictly increasing with strictly positive partial derivatives on R.

(iii) Utility $u(\cdot)$ is strongly quasi-concave on R:

$$\xi' \frac{\partial^2 u(x)}{\partial x \partial x'} \xi < 0,$$

for every $\xi \in \mathbb{R}^2$ such that $\xi \neq 0$ and $\xi' \frac{\partial u(x)}{\partial x} = 0$, at each $x \in R$.

(iv) For each $v \in \mathbb{R}$ such that $v \neq u(0)$, $G(v)$ is contained in R.

Assumption 1.6(i) assumes that preferences are smooth (see Proposition 2.3.9 in Mas-Colell, 1985). Assumption 1.6(ii) says that *more is strictly better*. Assumptions 1.6(i) and 1.6(ii) imply that $G(v)$ is an *indifference curve*, and twice-continuously-differentiable, if non-empty.[4] Assumption 1.6(iii) says that this indifference curve is strictly convex. Assumption 1.6(iii) implies strict quasi-concavity, a common assumption that requires the upper contour sets of the utility function to be strictly convex (Ginsberg, 1973). In other words, strong quasi-concavity implies that *averages are strictly better than extremes*. The converse does not hold because strict quasi-concavity allows for indifference curves to have zero curvature on nowhere-dense sets. Assumption 1.6(iv) says that the boundary of $\bar{R}$ is undesirable. Intuitively, the household prefers bundles with strictly positive quantities of both food and non-food, over bundles that do not. Assumption 1.6 is standard (see pages 415 to 416 in Katzner, 1968 and Sections 11 and 12 in Barten and Böhm, 1993).

Example 1.8. Suppose that the household has a Stone-Geary utility function, defined by: $u(x) = \sqrt{x_1 x_2}$, for every $x \in \bar{R}$. This utility function satisfies Assumption 1.6: First, note that, its first-order partial derivatives are well-defined and strictly positive:

$$\frac{\partial u(x)}{\partial x_1} = \sqrt{\frac{x_2}{4x_1}} > 0 \quad \text{and} \quad \frac{\partial u(x)}{\partial x_2} = \sqrt{\frac{x_1}{4x_2}} > 0, \tag{1.2.17}$$

at every $x \in R$. Moreover, the condition in Assumption 1.6(iii) equals:

$$\xi' \frac{\partial^2 u(x)}{\partial x \partial x'} \xi = -\frac{\xi_1^2}{4}\sqrt{\frac{x_2}{x_1^3}} - \frac{\xi_2^2}{4}\sqrt{\frac{x_1}{x_2^3}} + \frac{\xi_1 \xi_2}{2}\sqrt{\frac{1}{x_1 x_2}}, \tag{1.2.18}$$

for every $x \in R$, where ξ_j denotes the j^{th} component of ξ. The condition $\xi' \frac{\partial u(x)}{\partial x} = 0$ is equivalent to $\xi_1 = -\xi_2 x_1 / x_2$. Therefore:

$$\xi' \frac{\partial^2 u(x)}{\partial x \partial x'} \xi = -\xi_2^2 \sqrt{\frac{x_1}{x_2^3}}, \tag{1.2.19}$$

for every $\xi \in \mathbb{R}^2$ such that $\xi \neq 0$ and $\xi' \frac{\partial u(x)}{\partial x} = 0$, at each $x \in R$. Since $\xi \neq 0$ and $\xi' \frac{\partial u(x)}{\partial x} = 0$ implies $\xi_2 \neq 0$, this expression is strictly negative. Last, note, for every $v > 0$, the indifference curve $G(v) = \{x \in \bar{R} : x_1 x_2 = v^2\} \subset R$ does not intersect the boundary of the budget set. △

[4] If empty, the value $v \in \mathbb{R}$ is outside the range of utility $u(\cdot)$, and $G(v)$ is not an indifference curve.

1.2.5 Demand

The household is faced with the problem of choosing a bundle $x \in \bar{R}$, and an amount of fraud $f \in [0, b(z)]$. I assume that the design z and parameter $\theta = (b, \pi, u)$ are fixed, and that the household maximizes utility $u(\cdot)$. To be precise, let $x_\theta(z)$ denote its *demand for goods*, let $f_\theta(z)$ denote its *demand for fraud*, and consider the following assumption:

Assumption 1.7.

(i) The design z and parameter θ are exogenous to the household.

(ii) The demands, $x_\theta(z)$ and $f_\theta(z)$, solve the following utility maximization problem:

$$\begin{gathered} \max_{(x,f)} u(x) \text{ subject to } 0 \leq x_1, \ 0 \leq x_2, \\ 0 \leq f \leq b(z), \ px_1 + x_2 \leq \big[y + \pi f\big] + \big[b(z) - f\big], \text{ and } x_2 \leq y + \pi f. \end{gathered} \tag{1.2.20}$$

Assumption 1.7(i) implies that utility $u(\cdot)$ is exogenous to the household, ruling out endogenous preferences that can be affected by, say, the policy (see Bowles, 1998, for a broad discussion of endogenous preferences, and Hastings et al., 2020, for a relevant analysis of the effect of benefits on the composition of purchased foods). Assumption 1.7(ii) implies that the household will commit fraud if it has an immediate incentive to do so. This assumption is a strong behavioural restriction. In practice, the household (i) might not be able to find a retailer to collude with, (ii) might have an aversion to fraud, or (iii) might solve a dynamic optimization problem that explicitly accounts for the negative consequences associated with getting caught. (If a household is caught, it can be disqualified from receiving benefits, but benefits are considered a right, meaning that a lot of evidence is needed to disqualify a household. Therefore, in practice, very few households are disqualified.) Recall, this model is intended to serve as a benchmark.

1.2.6 The Budget Set

The utility maximization problem in (1.2.20) is somewhat complicated. We can simplify this maximization problem by using the fact that fraud f does not enter the objective function. But, in order to simplify this problem, we first need to characterize the set of all bundles in $\bar{R}$ that satisfy the budget constraints in (1.2.14), for *some* $f \in [0, b(z)]$.

Consider the following result:

Lemma 1.1. There exists $f \in [0, b(z)]$ that satisfies (1.2.14) given (z, b, π) if, and only if:

$$px_1 + x_2 \leq y + b(z) \text{ and } x_2 \leq y + \big[b(z) - px_1\big]\pi. \tag{1.2.21}$$

Proof. See Appendix 1.B.2. $\square$

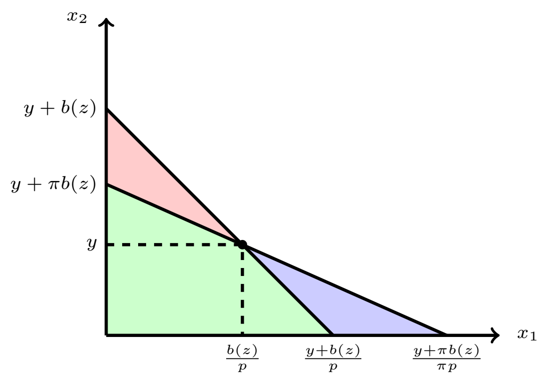

Figure 1.2. **The Budget Set.** The red and green regions characterize the subset of $\bar{R}$ for which $px_1 + x_2 \leq y + b(z)$. The blue and green regions characterize the subset of $\bar{R}$ for which $x_2 \leq y + [b(z) - px_1]\pi$. The budget set $B(z, b, \pi)$ is the green region.

Lemma 1.1 says that the household can afford a bundle $x \in \bar{R}$, with *some* amount of fraud $f \in [0, b(z)]$ if, and only if, (i) the entire bundle costs less than income y and benefits $b(z)$, and (ii) after scaling the price of food p by the discount factor π, the entire bundle costs less than income y and scaled benefits $\pi b(z)$. From now on, I let:

$$B(z, b, \pi) = \big\{x \in \bar{R} : (1.2.21) \text{ holds given } (z, b, \pi)\big\}, \tag{1.2.22}$$

denote the household's *budget set*—that is, the set of bundles in $\bar{R}$ that the household can afford with fraud given (z, b, π). I illustrate the form of this set in Figure 1.2. The household's budget set is kinked if, and only if, $b(z) > 0$. Kinked budget sets are common in applied economic settings—for instance, piecewise tax schedules, wholesale pricing, and two-part tariffs can all produce kinks (see Chapter 1 in Deaton and Muellbauer, 1980b, and Chapter 2.D in Mas-Colell et al., 1995, for some textbook examples, Blomquist et al., 2015, for a more recent application with taxable income, as well as Landais, 2015, for a more recent application with employment insurance). There are two limiting cases that are not allowed under Assumption 1.5: $\pi = 0$ and $\pi = 1$. In the first limiting case, the household cannot exchange its benefits for cash and the budget constraints in (1.2.21) reduce to the standard constraints in (1.2.10). In the second limiting case, there is no cost to an exchange (making benefits as good as cash) and the budget constraints in (1.2.21) reduce to $px_1 + x_2 \leq y + b(z)$, as expected.

The budget set has a number of important properties:

Lemma 1.2. The budget set $B(z, b, \pi)$ is non-empty, compact, and convex.

Proof. See Appendix 1.B.3. □

Lemma 1.2 implies that (i) the household can always afford a bundle (guaranteeing that it never faces a trivial decision), (ii) it can always afford the "limit" of affordable bundles, (iii) it can only afford finite-valued bundles, and (iv) it can always afford the convex combination of affordable bundles.

Example 1.7 (Continued). Suppose that the household has \$50 in income y and \$100 in benefits $b(z)$, and that the discount factor π is 0.5. It can afford a bundle $x \in \bar{R}$ if, and only if:

$$px_1 + x_2 \leq 150 \text{ and } 0.5px_1 + x_2 \leq 100. \tag{1.2.23}$$

△

1.2.7 Properties of Demand

We can now use the budget set to rewrite the utility maximization problem in (1.2.20), and characterize the properties of the demand functions, $x_\theta(\cdot)$ and $f_\theta(\cdot)$, given $\theta = (b, \pi, u)$.

Proposition 1.2. Under Assumptions 1.1 to 1.3 and 1.5 to 1.7, demand $x_\theta(z)$ solves:

$$\max_x \ u(x) \text{ subject to } x \in B(z, b, \pi). \tag{1.2.24}$$

Proof. This result follows directly from Lemma 1.1, and the fact that f does not enter the objective function in the utility maximization problem in (1.2.20). □

Proposition 1.2 implies that, if we are interested in demand $x_\theta(\cdot)$, then, instead of solving the utility maximization problem in (1.2.20) for $x_\theta(\cdot)$ and $f_\theta(\cdot)$, we can solve the well-behaved optimization problem in (1.2.24). Useful properties of demand $x_\theta(\cdot)$ follow from the fact that we are now simply maximizing $u(\cdot)$ over a non-empty, compact, and convex set.

Let $g(\cdot; z, b, \pi)$ denote the boundary of the budget set, defined by:

$$x_2 = g(x_1; z, b, \pi) \equiv \min\left\{y + b(z) - px_1, \ y + \left[b(z) - px_1\right]\pi\right\}, \tag{1.2.25}$$

for every $0 \leq x_1 \leq \frac{y+b(z)}{p}$. For every quantity of food x_1 in this range, $g(\cdot; z, b, \pi)$ outputs the largest quantity of non-food x_2 that the household can afford given (z, b, π). Furthermore, when $x_\theta(\cdot)$ is single-valued, I let $x_{\theta,j}(z)$ denote the j^{th} component of $x_\theta(z)$.

Proposition 1.3. Under Assumptions 1.1 to 1.3 and 1.5 to 1.7:

(i) Demand $x_\theta(\cdot)$ is well-defined, single-valued, and strictly positive, given θ, on $\mathbb{R}^2_{++}$.

(ii) Budget exhaustion: $x_{\theta,2}(z) = g(x_{\theta,1}(z); z, b, \pi)$.

Proof. See Appendix 1.B.4. □

Proposition 1.3(i) implies that $x_\theta(\cdot)$ is a positive function. Proposition 1.3(ii)—a variation of *Walras' Law* (Walras, 1874)—says that the household spends all of its income y, and spends or exchanges (and subsequently spends) all of its benefits $b(z)$.

Corollary 1.1. Under Assumptions 1.1 to 1.3 and 1.5 to 1.7:

$$f_\theta(z) = \big[b(z) - px_{\theta,1}(z)\big]^+, \tag{1.2.26}$$

for every $z \in \mathbb{R}^2_{++}$, and demand for fraud $f_\theta(\cdot)$ is well-defined and single-valued on $\mathbb{R}^2_{++}$.

Proof. See Appendix 1.B.5. □

Intuitively, if the household does not spend all of its benefits $b(z)$ on food, it will have $b(z) - px_{\theta,1}(z)$ in benefits left to exchange for cash. Corollary 1.1 implies that $f_\theta(\cdot)$ is a non-negative function, outputing this amount, whenever it is positive. The form in (1.2.26) directly implies that the chosen amount of fraud $f_\theta(z)$ can be deduced from only the knowledge of the benefit amount $b(z)$ and food expenditure $px_{\theta,1}(z)$.

1.2.8 The Regimes

In the absence of benefits, the utility maximization problem in (1.2.24) reduces to the *standard utility maximization problem* in economics under a linear budget constraint:

$$\max_x \; u(x) \;\text{ subject to }\; 0 \le x_1, \;\; 0 \le x_2, \;\text{ and }\; px_1 + x_2 \le y. \tag{1.2.27}$$

Let $x_u(z)$ denote the household's *standard demand for goods* given z and $u(\cdot)$—that is, the subset of bundles in $\bar{R}$ that solve the optimization problem in (1.2.27) given z and $u(\cdot)$. Because (1.2.27) is a special case of (1.2.24), the properties of $x_\theta(\cdot)$ translate into properties of $x_u(\cdot)$: Under Assumption 1.6, $x_u(\cdot)$ is well-defined, single-valued, strictly positive, and it satisfies Walras' law. It is also known that, under Assumption 1.6, the implicit function theorem implies that $x_u(\cdot)$ is continuously-differentiable on $\mathbb{R}^2_{++}$ (see Section 2.2.2 in Chapter 2 for a discussion and a proof in the case of two goods). If we were to replace strong quasi-concavity with strict quasi-concavity, $x_u(\cdot)$ would be continuously-differentiable on a dense open subset, but not necessarily everywhere (see Katzner, 1968). This standard utility maximization problem has been studied in detail (see Chapter 3.D in Mas-Colell et al., 1995). When $x_u(\cdot)$ is single-valued, I let $x_{u,j}(z)$ denote the j^{th} component of $x_u(z)$.

Standard demand $x_u(\cdot)$ can be used to construct a closed-form expression for demand $x_\theta(\cdot)$:

Proposition 1.4. Define $\phi(z) \equiv y + \pi b(z)$. Under Assumptions 1.1 to 1.3 and 1.5 to 1.7:

$$x_\theta(z) = \begin{cases} x_u(\psi(z), p), & \text{if } \frac{b(z)}{p} < x_{u,1}(\psi(z), p), \\ \left(\frac{b(z)}{p}, y\right)', & \text{if } x_{u,1}(\psi(z), p) \le \frac{b(z)}{p} \le x_{u,1}(\phi(z), \pi p), \\ x_u(\phi(z), \pi p), & \text{if } \frac{b(z)}{p} > x_{u,1}(\phi(z), \pi p), \end{cases} \tag{1.2.28}$$

given θ on $\mathbb{R}^2_{++}$.

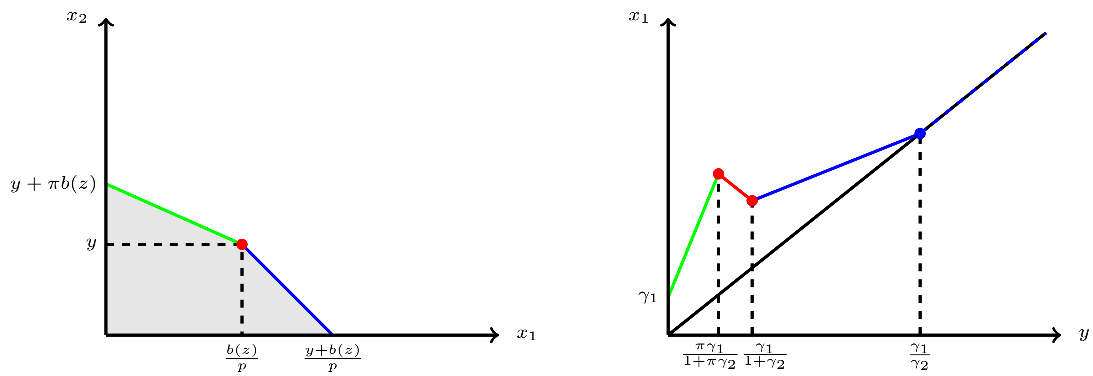

Figure 1.3. **The Regimes.** On the left, I illustrate $B(z, b, \pi)$ when $b(z) > 0$. On the right, I illustrate demand for food in (1.2.29) given $p = \frac{1}{2}$. Demand $x_\theta(\cdot)$ is in the first regime if it is on a blue segment, the second regime if it is on a red segment, and the third regime if it is on a green segment. On the right, the black line denotes standard food demand $x_{u,1}(\cdot)$.

Proof. See Appendix 1.B.6. $\square$

There are *three distinct regimes*: In the first regime, defined by $b(z)/p < x_{u,1}(\psi(z), p)$, the household buys more than $b(z)/p$ units of food; in the second regime, defined by $x_{u,1}(\psi(z), p) \leq b(z)/p \leq x_{u,1}(\phi(z), \pi p)$, it buys exactly $b(z)/p$ units of food; in the third regime, defined by $b(z)/p > x_{u,1}(\phi(z), \pi p)$, it buys less than $b(z)/p$ units of food (see the left of Figure 1.3 for an illustration). Of course, there is a *fourth regime*, defined by $b(z) = 0$, but this regime is a subset of the first regime. Now, recall that there are two limiting cases that are not considered under Assumption 1.5: $\pi = 0$ and $\pi = 1$. Intuitively, because $x_{u,1}(\phi(z), \pi p)$ gets large as the discount factor π gets arbitrarily close to 0, the third regime is empty in the first limiting case. In the second limiting case, we obtain: $x_\theta(z) = x_u(\psi(z), p)$.

Example 1.8 (Continued)**.** Suppose that the household has a Stone-Geary utility function, defined by: $u(x) = \sqrt{x_1 x_2}$, for every $x \in \bar{R}$. Under this specification, standard demand has the form: $x_u(z) = \left(\frac{y}{2p}, \frac{y}{2}\right)'$, for each $z \in \mathbb{R}^2_{++}$. Further suppose that the policy $b(\cdot)$ has the form with fixed prices in (1.2.6) subject to $\gamma_2 < 1$. Proposition 4 yields:

$$x_\theta(z) = \begin{cases} \left(\frac{y}{2p}, \frac{y}{2}\right)', & \text{if } y > \frac{\gamma_1}{\gamma_2}, \\ \left(\frac{\gamma_1 + y(1-\gamma_2)}{2p}, \frac{\gamma_1 + y(1-\gamma_2)}{2}\right)', & \text{if } \frac{\gamma_1}{1+\gamma_2} < y \leq \frac{\gamma_1}{\gamma_2}, \\ \left(\frac{\gamma_1 - \gamma_2 y}{p}, y\right)', & \text{if } \frac{\pi\gamma_1}{1+\pi\gamma_2} \leq y \leq \frac{\gamma_1}{1+\gamma_2}, \\ \left(\frac{\pi\gamma_1 + y(1-\pi\gamma_2)}{2\pi p}, \frac{\pi\gamma_1 + y(1-\pi\gamma_2)}{2}\right)', & \text{if } y < \frac{\pi\gamma_1}{1+\pi\gamma_2}. \end{cases} \tag{1.2.29}$$

There are *four cases*: In the first case, defined by $y > \gamma_1/\gamma_2$, the household does not receive any benefits. The remaining three cases characterize the first, second, and third regimes, in order, given $b(z) > 0$. The household's demand for food $x_{\theta,1}(\cdot)$ is strictly increasing in income y in the first and third regimes, and strictly decreasing in income y in the second regime (see the right of Figure 1.3). Under Assumptions

1.1 to 1.7, we obtain this result whenever food is a normal good—specifically, whenever standard food demand $x_{u,1}(\cdot)$ is strictly increasing in income y, given p. Moreover, we can use Corollary 1.1 to find the household's demand for fraud:

$$f_\theta(z) = \left[\frac{\pi\gamma_1 - (1+\pi\gamma_2)y}{2\pi}\right]^+, \tag{1.2.30}$$

for every $z \in \mathbb{R}^2_{++}$. This function is strictly decreasing in income y when it is strictly positive. Again, this result does not depend on the chosen policy $b(\cdot)$, but on the fact that food is a normal good. This result implies that, if food is a normal good, then we can reduce demand for fraud $f_\theta(\cdot)$ with an in-cash transfer. Of course, if implemented without care, this type of transfer can distort the ranking of eligible households of the same size with respect to purchasing power in the population. Any reasonable transfer scheme must ensure that total income $\psi(\cdot)$ is strictly increasing in income y. $\triangle$

Remark 1.2. Let $R_{\theta,j}$ denote the subset of $z \in \mathbb{R}^2_{++}$ at which demand $x_\theta(\cdot)$ is in the j^{th} regime. Since, under Assumptions 1.1 to 1.7, the policy $b(\cdot)$ and standard demand $x_u(\cdot)$ are continuous (see Propositions 1.1 and 1.3), the form of demand $x_\theta(\cdot)$ in Proposition 1.4 implies:

(i) $R_{\theta,1}$ and $R_{\theta,2}$ have a shared boundary.

(ii) $R_{\theta,2}$ and $R_{\theta,3}$ have a shared boundary.

(iii) $R_{\theta,1}$ and $R_{\theta,3}$ do not have a shared boundary.

Note that, Assumptions 1.1 to 1.3 and 1.5 to 1.7 are not sufficient for this property to hold: If the policy $b(\cdot)$ experiences a large jump when income y reaches the threshold $c(p)$, demand $x_\theta(\cdot)$ can go straight from the first regime to the third. Under Assumptions 1.1 to 1.7, the boundaries between the regimes are characterized by the "upper" and "lower" boundaries of the second regime, where these boundaries are curves in $\mathbb{R}^2_{++}$.

Example 1.8 (Continued)**.** Figure 1.4 illustrates Remark 1.2 using the regimes associated with the demand function $x_\theta(\cdot)$ in (1.2.29) given the policy $b(\cdot)$ in Example 1.2. Recall, this policy is continuous if $\gamma_3 = \gamma_1/\gamma_2$. I illustrate the sets $R_{\theta,j}$ given $\gamma_3 = \gamma_1/\gamma_2$ on the left in Figure 1.4. In this figure, we can see that, the property in Remark 1.2 holds as long as γ_3 is strictly smaller than $\frac{\pi\gamma_1}{1+\pi\gamma_2}$. I illustrate the sets $R_{\theta,j}$ given $\gamma_3 = \frac{\pi\gamma_1}{1+\pi\gamma_2}$ on the right in Figure 1.4. In this case, the first and third regimes have a shared boundary, and the second regime is empty. Intuitively, since only the poorest households have an incentive to commit fraud, if the threshold γ_3 is extremely small, as in this figure, the household will commit fraud as soon as it is eligible to receive positive benefits. $\triangle$

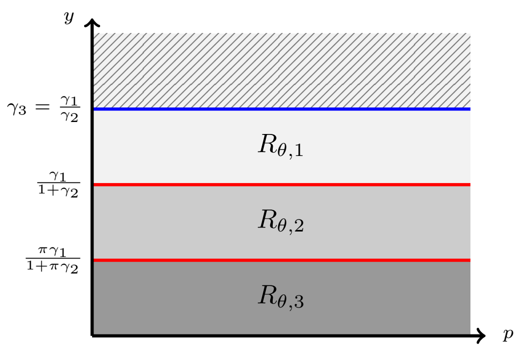

Figure 1.4. **Adjacent Regimes.** These figures illustrate the regimes associated with $x_\theta(\cdot)$ in (1.2.29) given $b(\cdot)$ in Example 1.2 for different values of γ_3. On the left, the regimes have the "order" in Remark 1.2. Hatched regions denote the sets of z on which $b(z) = 0$. Red lines denote boundaries of regimes. Blue lines denote sets of z on which $y = c(p)$.

1.2.9 Expenditure with Benefit Fraud

Although, in practice, we may not observe income y, we usually observe food expenditure $e_{\theta,1}(z)$, non-food expenditure $e_{\theta,2}(z)$, and total expenditure $e_\theta(z)$, related by the equation:

$$e_\theta(z) = e_{\theta,1}(z) + e_{\theta,2}(z) = px_{\theta,1}(z) + x_{\theta,2}(z). \tag{1.2.31}$$

The following result implies that we can use these expenditures to characterize the regimes:

Proposition 1.5. Under Assumptions 1.1 to 1.3 and 1.5 to 1.7:

(i) $z \in R_{\theta,1}$ if, and only if, $e_{\theta,1}(z) > b(z)$, $e_{\theta,2}(z) < y$, and $e_\theta(z) = y + b(z)$.

(ii) $z \in R_{\theta,2}$ if, and only if, $e_{\theta,1}(z) = b(z)$, $e_{\theta,2}(z) = y$, and $e_\theta(z) = y + b(z)$.

(iii) $z \in R_{\theta,3}$ if, and only if, $e_{\theta,1}(z) < b(z)$, $e_{\theta,2}(z) > y$, and $y + \pi b(z) < e_\theta(z) < y + b(z)$.

Proof. See Appendix 1.B.7. □

In the first regime, the household spends more than $b(z)$ on food, less than y on non-food, and exactly $y + b(z)$ in total; in the second regime, it spends exactly $b(z)$ on food, exactly y on non-food, and exactly $y + b(z)$ in total; in the third regime, it spends less than $b(z)$ on food, more than y on non-food, and an amount more than $y + \pi b(z)$, but less than $y + b(z)$ in total.

Example 1.8 (Continued)**.** Suppose that the household has a Stone-Geary utility function, defined by: $u(x) = \sqrt{x_1 x_2}$, for every $x \in \bar{R}$. Further suppose that the policy $b(\cdot)$ has the form with fixed prices in (1.2.6) subject to $\gamma_2 < 1$. Under these specifications, demand $x_\theta(\cdot)$ has the form in (1.2.29). Thus, total expenditure $e_\theta(\cdot)$ has the following form:

$$e_\theta(z) = \begin{cases} y, & \text{if } y > \frac{\gamma_1}{\gamma_2}, \\ \gamma_1 + (1-\gamma_2)y, & \text{if } \frac{\pi\gamma_1}{1+\pi\gamma_2} \leq y \leq \frac{\gamma_1}{\gamma_2}, \\ \frac{(1+\pi)(\pi\gamma_1 + (1-\pi\gamma_2)y)}{2\pi}, & \text{if } y < \frac{\pi\gamma_1}{1+\pi\gamma_2}. \end{cases} \tag{1.2.32}$$

There are *three cases*: the first case coincides with the first case in (1.2.29); the second case coincides with the second and third cases in (1.2.29); the third case coincides with the fourth case in (1.2.29). The second and third cases in (1.2.29) are grouped together here, because, in each of these cases, $e_\theta(z) = y + b(z)$ and $b(z) > 0$. In the third case:

$$y + \pi b(z) = \pi\gamma_1 + (1-\pi\gamma_2)y < e_\theta(z) < \gamma_1 + (1-\gamma_2) = y + b(z), \tag{1.2.33}$$

implying that the inequalities in Proposition 1.5(iii) hold. The first inequality in (1.2.33) holds because $\pi < 1$, and the second inequality in (1.2.33) holds whenever $y < \frac{\pi\gamma_1}{1+\pi\gamma_2}$. $\triangle$

1.2.10 Income Effects

We will sometimes need more structure for identification or non-parametric estimation. In this section, I assume that food and non-food are *normal goods*, and discuss the implication of this assumption on the effect of a change in income y on demand $x_\theta(\cdot)$.

Assumption N. Standard demand $x_u(\cdot)$ is strictly increasing in y, given p, on $\mathbb{R}_{++}$.

Without Assumption N, there exists a good $j = 1, 2$ where standard demand $x_{u,j}(\cdot)$ is increasing in y, at each $z \in \mathbb{R}^2_{++}$, but this property does not necessarily hold for this good at every $z \in \mathbb{R}^2_{++}$, and it does not necessarily hold for both goods. In general, it is not reasonable to assume that a good is normal when it is extremely disaggregate, but can be reasonable for some aggregate goods, especially when they aggregate a lot of goods of the same type with all levels of quality (see the discussion of normal goods on page 25 of Mas-Colell et al., 1995). There are substitutes for goods classified as "food"—for instance, hot meals prepared for immediate consumption are classified as "non-food" (see Appendix 1.A.8)—but goods classified as "food" are primarily *staples* (e.g., bread, vegetables, milk, meat), making it difficult to find a natural substitute. A large change in income could lead to a household substituting goods classified as "food" for prepared meals, but, in my application, I restrict the sample to relatively poor households, making this possibility less of a concern. Assumption N holds under some well-known conditions, such as *homotheticity*—that is, it is sufficient for utility $u(\cdot)$ to be homogeneous of degree one (see Leroux, 1987, for an alternative sufficient condition concerning the partial derivatives of utility). Assumption N is helpful because, if we consider it together with Assumption 1.4, then, as income goes up, we will pass between regimes, but never leave and then re-enter a regime (see Figure 1.4). In most cases where

Assumption N is needed, we will actually only need food to be a normal good; I will make it clear when we also need non-food to be a normal good.

Proposition 1.6. Under Assumptions 1.1 to 1.7, and N, demand for food $x_{\theta,1}(\cdot)$ is strictly increasing in y on $R_{\theta,1} \cup R_{\theta,3}$, and non-increasing in y on $R_{\theta,2}$. Moreover, demand for non-food $x_{\theta,2}(\cdot)$ is strictly increasing in y on $\mathbb{R}^2_{++}$.

Proof. This result follows from Proposition 1.4, Assumption N, and the fact that, under Assumption 1.4, $\psi(z) = y + b(z)$ and $\phi(z) = y + \pi b(z)$ are strictly increasing in y. $\square$

Proposition 1.6 says that demand for food $x_{\theta,1}(\cdot)$ is strictly increasing in income y if, and only if, it is in the first or third regimes, and that demand for non-food $x_{\theta,2}(\cdot)$ is strictly increasing everywhere. The first property can be seen on the right of Figure 1.3: If the household is extremely poor, it will consume on the green segment; as income increases, demand for food increases; when we hit the red segment, demand for food begins to decrease; when we hit the blue segment, demand for food begins to increase again, and continues indefinitely.

1.2.11 Differentiability of Demand

Assumption N provides the structure we need to make precise statements about the differentiability of demand $x_\theta(\cdot)$ with respect to income y. Consider the following result:

Proposition 1.7. Under Assumptions 1.1 to 1.7, and N, demand $x_\theta(\cdot)$ is continuously-differentiable at $z \in \mathbb{R}^2_{++}$ with respect to y, given p, if it is not on the boundary of a regime, and $y \neq c(p)$. Demand $x_\theta(\cdot)$ is not differentiable on the boundary of any regime, and, if the policy $b(\cdot)$ is not differentiable at some $z \in \mathbb{R}^2_{++}$, then neither is demand $x_\theta(\cdot)$.

Proof. See Appendix 1.B.8. $\square$

Proposition 1.7 ensures that there are no "jumps" in the rate at which demand $x_\theta(\cdot)$ changes from a change in income y if demand $x_\theta(\cdot)$ is not on the boundary of a regime and income y is not at the threshold $c(p)$. Proposition 1.7 implies that demand $x_\theta(\cdot)$ is not differentiable along two curves in $\mathbb{R}^2_{++}$: the upper and lower boundaries of the second regime (see Remark 1.2). Moreover, it implies that demand $x_\theta(\cdot)$ is not differentiable when y hits the threshold $c(p)$ if, and only if, the policy $b(\cdot)$ is not differentiable at this point, producing a third curve on which demand $x_\theta(\cdot)$ may not be differentiable. To illustrate, on the left of Figure 1.4, demand $x_\theta(\cdot)$ is continuously-differentiable everywhere except the three coloured curves, in which the red curves denote the boundaries of the regimes, and the blue curve denotes the curve characterized by $y = c(p)$. Proposition 1.7 relies on the discount factor π being strictly smaller than 1—in the limiting case for which $\pi = 1$, demand $x_\theta(z)$ coincides with standard demand $x_u(\psi(z), p)$,

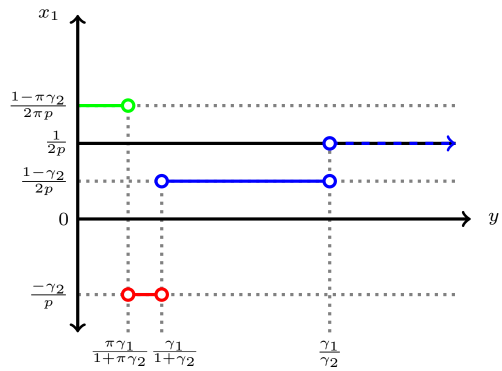

Figure 1.5. **Derivative of Demand.** This figure illustrates the partial derivative of demand for food $x_{\theta,1}(\cdot)$ in (1.2.29) with respect to income y. Demand $x_\theta(\cdot)$ is in the first regime if the partial derivative is on the blue segment, the second regime if it is on the red segment, and the third regime if it is on the green segment. On the right, the black line is the partial derivative of standard food demand with respect to income y.

ensuring that it is continuously-differentiable at $z \in \mathbb{R}^2_{++}$ if, and only if, the policy $b(\cdot)$ is continuously-differentiable at z. Differentiability is not needed for most of the results that follow, but it makes the argument in the next section easier, and it makes non-parametric estimation possible.

Example 1.8 (Continued)**.** Suppose that the household has a Stone-Geary utility function, defined by: $u(x) = \sqrt{x_1 x_2}$, for every $x \in \bar{R}$. Further suppose that the policy $b(\cdot)$ has the form with fixed prices in (1.2.6) subject to $\gamma_2 < 1$. For every $z \in \mathbb{R}^2_{++}$ with y not equal to $\frac{\gamma_1}{\gamma_2}$, $\frac{\gamma_1}{1+\gamma_2}$, or $\frac{\pi\gamma_1}{1+\pi\gamma_2}$, we obtain:

$$\frac{\partial x_\theta(z)}{\partial y} = \begin{cases} \left(\frac{1}{2p}, \frac{1}{2}\right)', & \text{if } y > \frac{\gamma_1}{\gamma_2}, \\ \left(\frac{1-\gamma_2}{2p}, \frac{1-\gamma_2}{2}\right)', & \text{if } \frac{\gamma_1}{1+\gamma_2} < y < \frac{\gamma_1}{\gamma_2}, \\ \left(\frac{-\gamma_2}{p}, 1\right)', & \text{if } \frac{\pi\gamma_1}{1+\pi\gamma_2} < y < \frac{\gamma_1}{1+\gamma_2}, \\ \left(\frac{1-\pi\gamma_2}{2\pi p}, \frac{1-\pi\gamma_2}{2}\right)', & \text{if } y < \frac{\pi\gamma_1}{1+\pi\gamma_2}. \end{cases} \tag{1.2.34}$$

Demand $x_\theta(\cdot)$ is, therefore, not differentiable at any $z \in \mathbb{R}^2_{++}$ with y equal to $\frac{\gamma_1}{\gamma_2}$, $\frac{\gamma_1}{1+\gamma_2}$, or $\frac{\pi\gamma_1}{1+\pi\gamma_2}$. The partial derivatives of demand $x_\theta(\cdot)$ in (1.2.34) are illustrated in Figure 1.5. △

1.2.12 Invertibility of Total Expenditure

Propositions 1.6 and 1.7 provide the foundation for a useful result: the invertibility of total expenditure $e_\theta(\cdot)$ in income y. Let $w = (e, p)$ denote a pair of expenditure and price.

Proposition 1.8. Under Assumptions 1.1 to 1.7, and N, there exists a function $y^*_\theta(\cdot)$ where $e_\theta(y^*_\theta(w), p) = e$ on the set of admissible values for the pair $w = (e, p)$ in $\mathbb{R}^2_{++}$. This function is continuous and strictly increasing in e on this set.

Proof. See Appendix 1.B.9. □

I refer to $y^*_\theta(\cdot)$ as the *pseudo-income function*. Intuitively, $y^*_\theta(e,p)$ denotes an amount of income y at which the household spends exactly e in total, given p. By Proposition 1.8, this amount is unique, continuous, and strictly increasing in total expenditure e. In the absence of benefits, the pseudo-income function $y^*_\theta(\cdot)$ coincides with total expenditure e, a direct implication of Walras' law. Once again, let us discuss the two limiting cases that are not considered under Assumption 1.5: $\pi = 0$ and $\pi = 1$. This discussion will help us understand what to expect for small and large values of the discount factor π, and help us compare with the standard model of consumption under a linear budget constraint in the absence of benefits and fraud. Recall, (i) in the first limiting case, the third regime is empty, and (ii) in the second limiting case: $x_\theta(z) = x_u(\psi(z), p)$. Therefore, in each limiting case, total expenditure $e_\theta(\cdot)$ coincides with total income $\psi(\cdot)$, so that the pseudo-income function $y^*_\theta(\cdot)$ coincides with the inverse of total income $\psi(\cdot)$ given p—specifically, it coincides with the amount of income y that produces $\psi(z) = e$.

Example 1.8 (Continued)**.** Suppose that the household has a Stone-Geary utility function, defined by: $u(x) = \sqrt{x_1 x_2}$, for every $x \in \bar{R}$. Further suppose that the policy $b(\cdot)$ has the form in (1.2.6) subject to $\gamma_2 < 1$. By inverting total expenditure $e_\theta(\cdot)$, we obtain:

$$y^*_\theta(w) = \begin{cases} e, & \text{if } e > \frac{\gamma_1}{\gamma_2}, \\ \frac{e-\gamma_1}{1-\gamma_2}, & \text{if } \frac{(1+\pi)\gamma_1}{1+\pi\gamma_2} \le e \le \frac{\gamma_1}{\gamma_2}, \\ \frac{2\pi e - \pi(1+\pi)\gamma_1}{(1+\pi)(1-\pi\gamma_2)}, & \text{if } \frac{(1+\pi)\gamma_1}{2} < e < \frac{(1+\pi)\gamma_1}{1+\pi\gamma_2}, \end{cases} \tag{1.2.35}$$

There are *three cases*: Each case coincides with the respective case in (1.2.32). In the first case, income y coincides with total expenditure e because the household does not get benefits; in the second case, income y is smaller than total expenditure e because the household is spending a positive amount of benefits; in the third case, income y is smaller than total expenditure e, but not by as much as in the second case, because the household "loses" a portion of its benefits when it exchanges them for cash. The lower bound on e in the third case follows from the fact that total expenditure $e_\theta(z)$ tends to the ratio $\frac{(1+\pi)\gamma_1}{2}$ as income y tends to zero, while holding all other variables constant. △

Proposition 1.8 is used in Section 1.3.3, where we observe total expenditure e, but not income y. When income y is not observed, it is natural to think of demand $x_\theta(\cdot)$ as a function of total expenditure e, instead of income y, but we can only think about it in this way if there exists a bijection $y^*_\theta(\cdot)$ between e and y, given p.[5] When this bijection exists, as in Proposition 1.8, I write:

$$b^*_\theta(w) \equiv b(y^*_\theta(w), p), \quad x^*_\theta(w) \equiv x_\theta(y^*_\theta(w), p), \quad f^*_\theta(w) \equiv f_\theta(y^*_\theta(w), p). \tag{1.2.36}$$

I refer to these objects as the *pseudo-policy, pseudo-demand, and pseudo-fraud functions*, respectively. These functions differ from their counterparts introduced in Sections 1.2.1 and 1.2.5 because they depend

[5]This type of reformulation is common in the revealed preference literature. In a standard framework (without benefits or fraud), this reformulation is trivial. In the current framework, we need Assumption N to ensure invertibility.

on total expenditure e, rather than income y. Since $y^*_\theta(\cdot)$ is a bijection given p, it is equivalent to have knowledge of $x_\theta(\cdot)$, or $y^*_\theta(\cdot)$ and $x^*_\theta(\cdot)$.

Under Assumptions 1.1 to 1.7, and N, pseudo-demand $x^*_\theta(\cdot)$ adopts many of the properties of demand $x_\theta(\cdot)$—it is well-defined, single-valued, strictly positive, and it satisfies a variant of Walras' law such that:

$$px^*_{\theta,1}(w) + x^*_{\theta,2}(w) = e, \tag{1.2.37}$$

at every admissible $w \in \mathbb{R}^2_{++}$, where $x^*_{\theta,j}(w)$ denotes the j^{th} component of $x^*_\theta(w)$. It is also continuously-differentiable at some admissible $w \in \mathbb{R}^2_{++}$ with respect to total expenditure e, given p, if it is not on the boundary of a regime, and not differentiable on the boundary of any regime. To be clear, if pseudo-demand $x^*_\theta(\cdot)$ is not on the boundary of a regime, then it is continuously-differentiable in total expenditure e, even if demand $x_\theta(\cdot)$ is not differentiable at $(y^*_\theta(w), p)'$. Intuitively, this lack of differentiability follows from the fact that pseudo-demand $x^*_{\theta,1}(\cdot)$ forms a *ridge* along the boundary of the third regime, and a *valley* along the boundary of the first regime. (This terminology is taken from differential geometry.) We can also use standard demand $x_u(\cdot)$ to construct a closed-form expression for pseudo-demand $x^*_\theta(\cdot)$, as in Section 1.2.8:

Corollary 1.2. Define total income evaluated at pseudo-income $\psi^*_\theta(w) \equiv y^*_\theta(w) + b^*_\theta(w)$ and the amount $\phi^*_\theta(w) \equiv y^*_\theta(w) + \pi b^*_\theta(w)$. Under Assumptions 1.1 to 1.7, and N:

$$x^*_\theta(w) = \begin{cases} x_u(\psi^*_\theta(w), p), & \text{if } \frac{b^*_\theta(w)}{p} < x_{u,1}(\psi^*_\theta(w), p), \\ \left(\frac{b^*_\theta(w)}{p}, y^*_\theta(w)\right)', & \text{if } x_{u,1}(\psi^*_\theta(w), p) \le \frac{b^*_\theta(w)}{p} \le x_{u,1}(\phi^*_\theta(w), \pi p), \\ x_u(\phi^*_\theta(w), \pi p), & \text{if } \frac{b^*_\theta(w)}{p} > x_{u,1}(\phi^*_\theta(w), \pi p), \end{cases} \tag{1.2.38}$$

given θ, at every admissible $w \in \mathbb{R}^2_{++}$.

Proof. This result follows from replacing income y with $y^*_\theta(w)$ in Proposition 1.4. □

Corollary 1.2 is an analogue of the result regarding the form of demand $x_\theta(\cdot)$ in Proposition 1.4 for pseudo-demand $x^*_\theta(\cdot)$. The regimes in Corollary 1.2 coincide with the regimes in Proposition 1.4, but the regimes are now associated with values of $w = (e, p)$, instead of $z = (y, p)$. Let $W_{\theta,j}$ denote the set of admissible $w \in \mathbb{R}^2_{++}$ at which pseudo-demand $x^*_\theta(\cdot)$ is in the j^{th} regime. Note, by definition, $w \in W_{\theta,j}$ if, and only if, $(y^*_\theta(w), p)' \in R_{\theta,j}$.

Corollary 1.3. Under Assumptions 1.1 to 1.7, and N, pseudo-demand for food $x^*_{\theta,1}(\cdot)$ is strictly increasing in e on $W_{\theta,1} \cup W_{\theta,3}$, and non-increasing in e on $W_{\theta,2}$. Moreover, pseudo-demand for non-food $x^*_{\theta,2}(\cdot)$ is strictly increasing in e at all admissible $w \in \mathbb{R}^2_{++}$.

Proof. This result follows from Proposition 1.6, the definition of pseudo-demand $x^*_\theta(\cdot)$ in (1.2.36), and the fact that, by Proposition 1.8, pseudo-income $y^*_\theta(\cdot)$ is strictly increasing in e. □

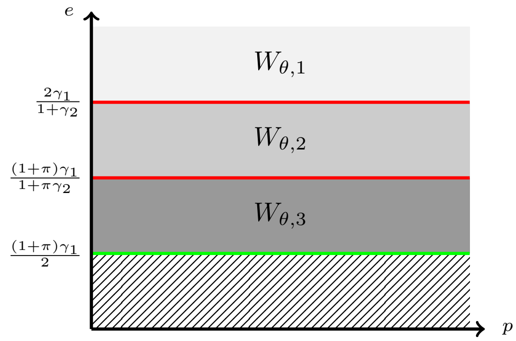

Figure 1.6. **The Regimes of Pseudo-Demand.** This figure illustrates the regimes associated with $x^*_\theta(\cdot)$ in (1.2.40) given the policy $b(\cdot)$ with fixed prices in (1.2.6). Red lines denote boundaries of regimes. Pairs $w = (e, p)$ in the hatched region are inadmissible.

Corollary 1.3 is an analogue of Proposition 1.6. I state this result formally because it is essential for some of the identification results in Section 1.3.3. Corollary 1.3 implies that we can use *expenditure effects* to characterize each of the regimes. In particular, we can use the curves on which pseudo-demand $x^*_\theta(\cdot)$ is not differentiable to partition the set of all admissible pairs $w \in \mathbb{R}^2_{++}$ into three subsets (which coincide with the regimes), then determine which set is which regime using Corollary 1.3 and what we know about the relationship between the regimes (see Section 1.3.3 for a rule that can be used to classify the elements of this partition).

Example 1.8 (Continued)**.** Suppose that the household has a Stone-Geary utility function, defined by: $u(x) = \sqrt{x_1 x_2}$, for every $x \in \bar{R}$. Further suppose that the policy $b(\cdot)$ has the form with fixed prices in (1.2.6) subject to $\gamma_2 < 1$. If we plug the pseudo-income function $y^*_\theta(\cdot)$ in (1.2.35) into the policy $b(\cdot)$, then:

$$b^*_\theta(w) = \begin{cases} 0, & \text{if } e > \frac{\gamma_1}{\gamma_2}, \\ \frac{\gamma_1 - \gamma_2 e}{1-\gamma_2}, & \text{if } \frac{(1+\pi)\gamma_1}{1+\pi\gamma_2} \le e \le \frac{\gamma_1}{\gamma_2}, \\ \frac{(1+\pi)\gamma_1 - 2\pi e \gamma_2}{(1+\pi)(1-\pi\gamma_2)}, & \text{if } \frac{(1+\pi)\gamma_1}{2} < e < \frac{(1+\pi)\gamma_1}{1+\pi\gamma_2}. \end{cases} \tag{1.2.39}$$

Each case above coincides with the respective case in (1.2.35). There are more cases in (1.2.39) than in the policy $b(\cdot)$ because the form of expenditure $e_\theta(\cdot)$ changes when demand $x_\theta(\cdot)$ enters the third regime.

Likewise, if we plug the pseudo-income function $y^*_\theta(\cdot)$ in (1.2.35) into the demand function in (1.2.29), then we obtain:

$$x^*_\theta(w) = \begin{cases} \left(\frac{e}{2p}, \frac{e}{2}\right)', & \text{if } e > \frac{2\gamma_1}{1+\gamma_2}, \\ \left(\frac{\gamma_1 - \gamma_2 e}{p(1-\gamma_2)}, \frac{e-\gamma_1}{1-\gamma_2}\right)', & \text{if } \frac{(1+\pi)\gamma_1}{1+\pi\gamma_2} \le e \le \frac{2\gamma_1}{1+\gamma_2}, \\ \left(\frac{e}{(1+\pi)p}, \frac{\pi e}{1+\pi}\right)', & \text{if } \frac{(1+\pi)\gamma_1}{2} < e < \frac{(1+\pi)\gamma_1}{1+\pi\gamma_2}, \end{cases} \tag{1.2.40}$$

The pseudo-demand function $x^*_\theta(\cdot)$ has different regimes than in (1.2.35) and (1.2.39) because, unlike $y^*_\theta(\cdot)$ and $b^*_\theta(\cdot)$, its form depends on the regime, not on whether benefits are zero or not—each case in

(1.2.40) coincides with the respective regime in Proposition 1.4. Clearly, pseudo-demand for food $x^*_{\theta,1}(\cdot)$ is strictly increasing in e in both the first and third regimes, and non-increasing in e in the third regime.

Finally, we can calculate:

$$f^*_\theta(w) = \left[b^*_\theta(w) - px^*_{\theta,1}(w)\right]^+ = \left[\frac{(1+\pi)\gamma_1 - (1+\pi\gamma_2)e}{(1+\pi)(1-\pi\gamma_2)}\right]^+. \tag{1.2.41}$$

While the form of the pseudo-fraud function $f^*_\theta(\cdot)$ differs from the form of the household's demand for fraud $f_\theta(\cdot)$ in (1.2.30), the pseudo-fraud function is non-increasing in total expenditure e, as demand for fraud $f_\theta(\cdot)$ is non-increasing in y. This result follows from the fact that $y^*_\theta(\cdot)$ is strictly increasing in e. △

Remark 1.3. It is worth noting that, if we were to construct a variant of the Slutsky matrix (see Section 2.F in Mas-Colell et al., 1995, for a broad discussion of the Slutsky matrix, as well as Allen, 1936, for a summary of work by Slutsky, 1915) by replacing standard demand $x_u(\cdot)$ with demand $x_\theta(\cdot)$ or pseudo-demand $x^*_\theta(\cdot)$, the resulting matrix would not necessarily be symmetric or negative semi-definite. Since these properties are used for integrability (see Samuelson, 1948, 1950, Hurwicz and Uzawa, 1971, Hosoya, 2020, and Section 1.3.4), a model without benefits might incorrectly reject rationality, even in the absence of fraud. Since only poor households receive benefits and only the poorest households have an incentive to consume outside of the first regime, this result can explain why some researchers find that, on average, poorer household are "less rational" (see, for example, Table 3 in Echenique et al., 2011).

1.3 Partial Non-Parametric Identification

Let us now consider the partial non-parametric identification of the functional parameter $\theta = (b, \pi, u)$, and some functions that depend on θ such as the demand for fraud $f_\theta(\cdot)$. Our objects of interest are latent, and do not depend on what we observe. In this section, I consider two observability assumptions: In the first assumption, we observe income y, the price p, and demand $x_\theta(z)$, at every $z = (y, p)$ in some set $\mathcal{Z}$; in the second assumption, we no longer observe income y, but we observe total expenditure e, the price p, and pseudo-demand $x^*_\theta(w)$, at every $w = (e, p)$ in some set $\mathcal{W}$. While the general form of the policy $b(\cdot)$ is known in practice, there are unobserved exclusions and deductions (see Appendices 1.A.4 and 1.A.6).[6]

1.3.1 Definition of Identification

Before moving forward, let us consider the definition of identification. This definition depends on what we observe. For exposition, I provide the definition of identification for θ under the first observability assumption: Assume that we observe the price p, income y, and demand $x_\theta(\cdot)$ over a set $\mathcal{Z}$ given θ. The

[6]If benefits are observed, as in the Panel Survey of Income Dynamics (see Appendix 1.E.1), the identification problem becomes much easier.

functional parameter θ is *partially identified* over a subset, say $\mathcal{Z}_0 \times \mathcal{X}_0$, if $x_\theta(z) = x_{\theta'}(z)$, for every $z \in \mathcal{Z}$, implies (i) $b(\cdot) = b'(\cdot)$ on $\mathcal{Z}_0$, (ii) $\pi = \pi'$, and (iii) $u(\cdot) = \varphi(\cdot) \circ u'(\cdot)$ on $\mathcal{X}_0$, for some strictly increasing function $\varphi(\cdot)$.[7] In this definition, the set $\mathcal{X}_0$ simply denotes an arbitrary subset of the consumption set $\bar{R}$. Instead of defining what it means for an object to be identified whenever a new object or assumption is introduced, I use the word "identified" colloquially to refer to this type of property. I focus on *partial identification* because the existence of a policy $b(\cdot)$ rules out *total identification* (see Appendix 1.C). This result has an important economic implication: We often need more information, or model structure, to deduce the effect of a change in a policy on objects of interest, like welfare or fraud. Sections 1.3.2 and 1.3.3 consider the identification of several latent functional parameters under each of the observability assumptions, starting with the case in which income y is observed. Section 1.3.4 considers, separately, the partial non-parametric identification of utility $u(\cdot)$ under the identification of standard demand $x_u(\cdot)$ over some well-behaved set.

1.3.2 Identification when Income is Observed

Here, I consider the non-parametric identification of the regimes, policy $b(\cdot)$, discount factor π, standard demand $x_u(\cdot)$, and demand for fraud $f_\theta(\cdot)$ in the case in which income y is observed.

Observability Assumption

Recall, $R_{\theta,j}$ denotes the subset of $z \in \mathbb{R}^2_{++}$ at which demand $x_\theta(\cdot)$ is in the j^{th} regime, as defined by the form in Proposition 1.4. Consider the following assumption:

Assumption B.1.

(i) We observe $z = (y, p)$ and $x_\theta(z)$, at every $z \in \mathcal{Z}$.

(ii) The set $\mathcal{Z}$ is open and path-connected.[8]

(iii) The observable component $R^*_{\theta,j} \equiv \mathcal{Z} \cap R_{\theta,j}$ is non-empty, for $j = 2, 3$.

We do not observe benefits $b(z)$. Assumption B.1(i) is a strong assumption because we rarely observe income y, and it is naïve for us to think that we perfectly observe demand $x_\theta(\cdot)$. Indeed, households (i) may not get benefits when eligible, and (ii) might misreport what they buy to hide fraud. In Section 1.3.3, I remove the assumption that income y is observed. Assumption B.1(ii) is mostly for simplicity; I will make it clear when it is needed. Assumption B.1(iii) says that our dataset has consumptions on the kink of the budget set, and fraudulent behaviours. Of course, this final restriction is needed because we cannot expect to identify, say, the discount factor π, without any fraud.

[7] This definition of identification is deterministic. It can be seen as a special case of identification in a stochastic framework. See Section 2 in Koopmans and Reiersøl (1950) and page 578 in Rothenberg (1971) for broad discussions, and Section 1.4, for a stochastic treatment.

[8] $\mathcal{Z}$ is path-connected if, for any $z, z' \in \mathcal{Z}$, there exists $h : [0, 1] \to \mathcal{Z}$ with $h(0) = z$ and $h(1) = z'$.

Remark 1.4. Since the intersection of a finite number of open sets is open, Assumption B.1 implies that $R^*_{\theta,1}$ and $R^*_{\theta,3}$ are open. The set $R^*_{\theta,2}$ is not open because $R_{\theta,2}$ is closed, but $R^*_{\theta,2}$ always contains an open subset that is dense in $R^*_{\theta,2}$, ensuring that the closure of its interior equals the closure of $R^*_{\theta,2}$. Therefore, if we can identify a continuous function over the interior of $R^*_{\theta,2}$, then it is identified over $R^*_{\theta,2}$. Moreover, each set $R^*_{\theta,j}$ can be "disconnected," even though, under Assumption B.1, $\mathcal{Z}$ is path-connected.

There are *four steps* to identification under this observability assumption: (i) identify the regimes, (ii) identify the discount factor π, (iii) identify the policy $b(\cdot)$, and (iv) identify standard demand $x_u(\cdot)$ and demand for fraud $f_\theta(\cdot)$. The first step lays the foundation for the steps that follow; the second step lays the foundation for the third; the third step lays the foundation for the fourth. As previously mentioned, the identification of utility $u(\cdot)$ is considered, separately, in Section 1.3.4.

Identification of the Regimes

Before we can identify any latent functions of interest, we need to first identify the regimes.

Theorem 1.1. Under Assumptions 1.1 to 1.3, 1.5 to 1.7, and B.1:

(i) $z \in R^*_{\theta,1}$ if, and only if, $e_{\theta,2}(z) < y$,

(ii) $z \in R^*_{\theta,2}$ if, and only if, $e_{\theta,2}(z) = y$,

(iii) $z \in R^*_{\theta,3}$ if, and only if, $e_{\theta,2}(z) > y$,

for each $z \in \mathcal{Z}$. Therefore, the sets $R^*_{\theta,j}$ are identified and the regimes are observable.

Proof. This result follows from Proposition 1.5, and the fact that non-food expenditure:

$$e_{\theta,2}(z) = x_{\theta,2}(z), \tag{1.3.1}$$

is observed, at every $z \in \mathcal{Z}$. □

Theorem 1.1 can be seen on the left in Figure 1.3: Since non-food expenditure $e_{\theta,2}(z)$ equals demand for non-food $x_{\theta,2}(z)$, non-food expenditure $e_{\theta,2}(z)$ is smaller than income y if demand $x_\theta(\cdot)$ is on the blue segment, equal to income y if demand $x_\theta(\cdot)$ is on the red segment, and larger than income y if demand $x_\theta(\cdot)$ is on the green segment. Since the household is committing fraud when demand $x_\theta(\cdot)$ is in the third regime, Theorem 1.1 implies that we can identify the set of $z \in \mathcal{Z}$ on which demand for fraud $f_\theta(\cdot)$ is positive.

Identification of the Discount Factor

We are now in a position to address the *second step*: the identification of the discount factor π. The following result uses our knowledge of the regimes and their boundaries:

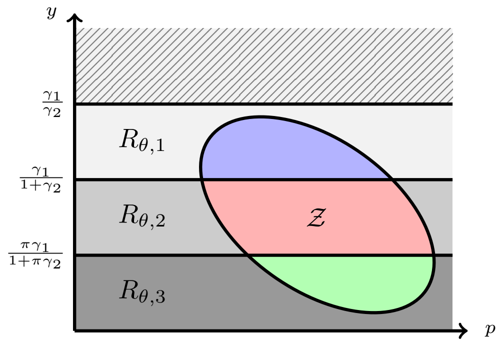

Figure 1.7. **The Observable Components of the Regimes.** This figure illustrates the regimes associated with $x_\theta(\cdot)$ in (1.2.29) given the policy $b(\cdot)$ with fixed prices in (1.2.6). The union of all coloured regions is $\mathcal{Z}$. The blue region is $R^*_{\theta,1}$; the red region is $R^*_{\theta,2}$; the green region is $R^*_{\theta,3}$. The hatched region denotes the set of z on which $b(\cdot)$ equals 0.

Theorem 1.2. Under Assumptions 1.1 to 1.7, N, and B.1:

(i) There exists $z_0 = (y_0, p_0) \in \mathcal{Z}$ at which demand x_θ is on the boundary of $R_{\theta,3}$.

(ii) The following equality holds:

$$\pi = \left(\frac{1}{p_0}\right) \frac{1 - \partial x_{\theta,2}(z_0)/\partial y_0^-}{\partial x_{\theta,1}(z_0)/\partial y_0^- - \partial x_{\theta,1}(z_0)/\partial y_0^+}, \tag{1.3.2}$$

for any observable $z_0 \in \mathcal{Z}$ on the boundary of $R_{\theta,3}$.

Therefore, the discount factor π is identified.

Proof. See Appendix 1.B.10. □

Theorem 1.2 says that we can deduce the rate at which the household can exchange its benefits $b(z)$ for cash. Theorem 1.2(i) follows from the properties of $\mathcal{Z}$ and the fact that we observe demand $x_\theta(\cdot)$ in the first and second regimes. Intuitively, demand $x_\theta(\cdot)$ cannot jump from the second regime to the third—it must cross their shared boundary. Theorem 1.2(ii) follows from the fact that the rate that demand $x_\theta(\cdot)$ changes with respect to income y changes at this boundary, and that this change in the rate depends on the discount factor π in a known way.

Example 1.8 (Continued)**.** Suppose that the household has a Stone-Geary utility function, defined by: $u(x) = \sqrt{x_1 x_2}$, for every $x \in \bar{R}$. Further suppose that the policy $b(\cdot)$ has the form with fixed prices in (1.2.6). Under these specifications, demand $x_\theta(\cdot)$ has the form in (1.2.29). Therefore, a design $z_0 \in \mathbb{R}^2_{++}$ is on the boundary of the third regime if, and only if, $y_0 = \frac{\pi\gamma_1}{1+\pi\gamma_2}$ (see Figure 1.4). Therefore, at every $z_0 \in \mathbb{R}^2_{++}$ on this boundary:

$$\frac{\partial x_{\theta,1}(z_0)}{\partial y_0^-} = \frac{1-\pi\gamma_2}{2\pi p_0}, \quad \frac{\partial e_{\theta,1}(z_0)}{\partial y_0^+} = -\frac{\gamma_2}{p_0}, \quad \frac{\partial e_{\theta,2}(z_0)}{\partial y_0^-} = \frac{1-\pi\gamma_2}{2}. \tag{1.3.3}$$

As expected, we obtain the following equality:

$$\left(\frac{1}{p_0}\right)\frac{1-\partial x_{\theta,2}(z_0)/\partial y_0^-}{\partial x_{\theta,1}(z_0)/\partial y_0^- - \partial x_{\theta,1}(z_0)/\partial y_0^+} = \frac{1-(1-\pi\gamma_2)/2}{(1-\pi\gamma_2)/2\pi+\gamma_2} = \pi. \tag{1.3.4}$$

△

Identification of the Benefit Policy

The identification of the discount factor π lays the foundation for the identification of the policy $b(\cdot)$. The role of the discount factor π is evident in the following theorem:

Theorem 1.3. Under Assumptions 1.1 to 1.7, N, and B.1:

(i) $z \in R^*_{\theta,1} \cup R^*_{\theta,2}$ implies $b(z) = e_\theta(z) - y$,

(ii) $z \in R^*_{\theta,3}$ implies $b(z) = \frac{1}{\pi}\big[\pi e_{\theta,1}(z) + e_{\theta,2}(z) - y\big]$,

for each $z \in \mathcal{Z}$. Therefore, the policy $b(\cdot)$ is identified over the set $\mathcal{Z}$.

Proof. Part (i) follows from Proposition 1.5, and part (ii) follows from budget exhaustion and the fact that $z \in R^*_{\theta,3}$ implies $g(x_{\theta,1}(z); z, b, \pi) = y + \big[b(z) - px_{\theta,1}(z)\big]\pi$. □

Theorem 1.3 says that we can identify the policy $b(\cdot)$ over the observable set $\mathcal{Z}$. In other words, we can deduce benefits $b(z)$ from household income y, the price p, and demand $x_\theta(z)$, even when fraud exists. Though the general form of the policy is known (see Appendix 1.A), our dataset might not have information on certain household characteristics that implicitly affect the policy $b(\cdot)$, making it difficult to use our knowledge of this form to explicitly calculate benefits $b(z)$. Here, Theorem 1.3 tells us that we can simply use budget exhaustion to deduce benefits $b(z)$. Since the equality in Theorem 1.3(ii) depends on the discount factor π, without Assumptions B.1(ii) and B.1(iii), we would not be able to identify the policy $b(\cdot)$ in the third regime.

Identification of Standard Demand and Demand for Fraud

The first three steps are needed for the *fourth step*: the identification of standard demand $x_u(\cdot)$ and demand for fraud $f_\theta(\cdot)$. Recall, $\psi(z)$ denotes total income $y+b(z)$, and $\phi(z)$ denotes the amount $y+\pi b(z)$. The next result follows from the closed-forms for the demands, $x_\theta(\cdot)$ and $f_\theta(\cdot)$, in (1.2.28) and (1.2.26).

Theorem 1.4. Under Assumptions 1.1 to 1.7, N, and B.1:

(i) Standard demand $x_u(\cdot)$ is identified over:

$$\left\{(\psi(z), p)' : z \in R^*_{\theta,1}\right\} \cup \left\{(\phi(z), \pi p)' : z \in R^*_{\theta,3}\right\}. \tag{1.3.5}$$

(ii) Demand for fraud $f_\theta(z) = [b(z) - px_{\theta,1}(z)]^+$ is identified over the set $\mathcal{Z}$.

Proof. See Corollary 1.1, Proposition 1.4, and Theorems 1.2 and 1.3. □

Theorem 1.4(i) says that we can identify part of standard demand $x_u(\cdot)$. Knowledge of standard demand $x_u(\cdot)$ is needed to evaluate the effect of a change in the policy $b(\cdot)$ on demand $x_\theta(\cdot)$ and demand for fraud $f_\theta(\cdot)$. It can also be used to recover utility $u(\cdot)$ (see Section 1.3.4). Theorem 1.4(i) follows from the fact that (i) in the first regime, demand $x_\theta(z)$ equals $x_u(\psi(z), p)$, and (ii) in the third regime, demand $x_\theta(z)$ equals $x_u(\phi(z), \pi p)$. Theorem 1.4(ii) says that we can identify the chosen amount of fraud $f_\theta(z)$, at every $z \in \mathcal{Z}$. If Assumption B.1(ii) or B.1(iii) is violated and we cannot identify the discount factor π, then we can only identify standard demand $x_u(\cdot)$ using the first regime, and we can only identify bounds for fraud $f_\theta(z)$.

1.3.3 Identification when Income is Not Observed

Up until now, I have assumed that we observe income y. I will now relax this assumption. As described in Section 1.2.12, when income y is not observed, it is common to think of demand, and other latent functions, as functions of total expenditure e, rather than income y. In this section, I consider the identification of the regimes, policy $b(\cdot)$, discount factor π, standard demand $x_u(\cdot)$, and demand for fraud $f_\theta(\cdot)$, when income y is not observed. Moreover, I will discuss the identification of some reduced-form objects: the pseudo-income function $y^*_\theta(\cdot)$, pseudo-policy function $b^*_\theta(\cdot)$, and pseudo-fraud function $f^*_\theta(\cdot)$. While these reduced-form objects do not provide additional information to the FNS, they let the econometrician immediately deduce the household's income y, benefits $b(z)$, and chosen amount of fraud $f_\theta(z)$, when income y is not observed. Appendix 1.D contains more on the identification of these reduced-form objects. Howevers, few objects are identified without additional model structure. In what follows, I focus on the identification of bounds, instead of imposing, say, strong parametric assumptions.

Observability Assumption

Recall, $W_{\theta,j}$ denotes the set of admissible $w \in \mathbb{R}^2_{++}$ at which pseudo-demand $x^*_\theta(\cdot)$ is in the j^{th} regime, as defined by the form in Corollary 1.2. Now, consider a new assumption:

Assumption B.2.

(i) We observe $w = (e, p)$ and $x^*_\theta(w)$, at every $w \in \mathcal{W}$.

(ii) The set $\mathcal{W}$ is open and path-connected.

(iii) The observable component $W^*_{\theta,j} \equiv \mathcal{W} \cap W_{\theta,j}$ is non-empty, for $j = 1, 2, 3$.

(iv) For each $w_0 \in \mathcal{W}$, there exists $w_1 \in W^*_{\theta,2}$ with $p_1 = p_0$.

Assumption B.2 does not assume that we observe income y, as in Assumption B.1, but it imposes some additional restrictions on $\mathcal{W}$: Assumption B.2(iii) says that *every regime* is non-empty; Assumption

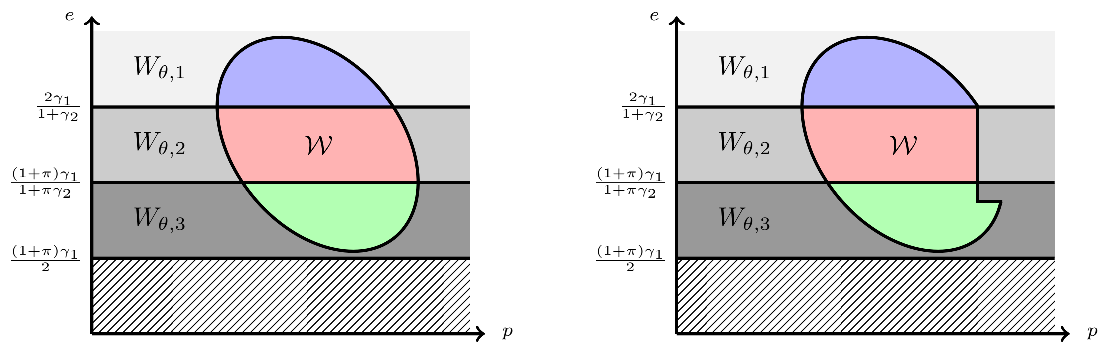

Figure 1.8. **Observable Set.** These figures illustrate the regimes associated with $x_\theta^*(\cdot)$ in (1.2.40) given the policy $b(\cdot)$ with fixed prices in (1.2.6). In each figure, the union of all coloured sets is $\mathcal{W}$, the blue region is $W_{\theta,1}^*$, the red region is $W_{\theta,2}^*$, and the green region is $W_{\theta,3}^*$. On the left, Assumption B.2 is satisfied. On the right, Assumption B.2(iv) is, in fact, violated because some pairs in $W_{\theta,3}^*$ are not directly below some pair in $W_{\theta,2}^*$.

B.2(iv) says that, for every observable price p, there is an observable amount e where (e, p) is in the second regime. Assumptions B.2(ii) to B.2(iv) are for simplicity; I will make it clear when they are needed.

When income y is observed, we can identify the regimes, use the boundary of the third regime to identify the discount factor π, use the discount factor π to identify the policy $b(\cdot)$, and use the policy $b(\cdot)$ to identify standard demand $x_u(\cdot)$ and demand for fraud $f_\theta(\cdot)$. Unfortunately, these steps are not feasible when income y is not observed—we can no longer identify the discount factor π using only our knowledge of the regimes.

The modified steps are as follows: (i) identify the regimes, (ii) identify bounds for the policy $b(\cdot)$, (iii) identify standard demand $x_u(\cdot)$, (iv) identify a bound for the discount factor π, and (v) identify bounds for the demand for fraud $f_\theta(\cdot)$. The first step lays the foundation for the steps that follow; the fourth step lays the foundation for the fifth.

Identification of the Regimes

In Section 1.3.2, we identified the regimes by comparing non-food expenditure $e_{\theta,2}(z)$ with income y. Since income y is no longer observed, we need a new way to identify the regimes. Consider a new result:

Theorem 1.5. Let $w_0 = (e_0, p_0) \in \mathcal{W}$ denote an observable pair, and let $w_1 = (e_1, p_1) \in W_{\theta,2}^*$ denote another observable pair that satisfies $p_1 = p_0$. Under Assumptions 1.1 to 1.7, N, and B.2, these pairs exist, and the following implications hold:

(i) If pseudo-demand for food $x_{\theta,1}^*(\cdot)$ is non-increasing in e at $w_0 \in \mathcal{W}$, then $w_0 \in W_{\theta,2}^*$.

(ii) If $w_0 \notin W_{\theta,2}^*$, then:

- $e_1 > e_0$ implies $w_0 \in W_{\theta,1}^*$;
- $e_1 < e_0$ implies $w_0 \in W_{\theta,3}^*$.

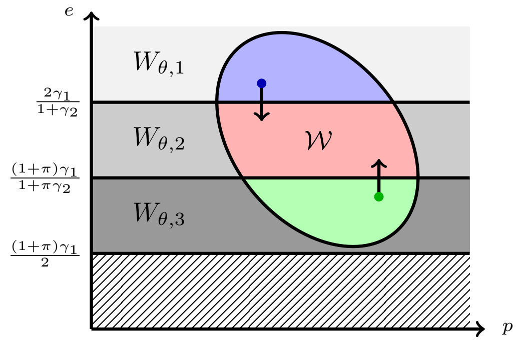

Figure 1.9. **Procedure in Theorem 1.5.** This figure illustrates the regimes associated with $x^*_\theta(\cdot)$ in (1.2.40) given the policy $b(\cdot)$ with fixed prices in (1.2.6). The union of all coloured sets is $\mathcal{W}$. The blue region is $W^*_{\theta,1}$, the red region is $W^*_{\theta,2}$, and the green region is $W^*_{\theta,3}$. After identifying $W^*_{\theta,2}$, we can classify the blue and green regions by looking for points that are above or below the red region, as shown by the blue and green nodes.

Therefore, the sets $W^*_{\theta,j}$ are identified and the regimes are observable.

Proof. See Corollary 1.3. □

Theorem 1.5 implies that we can identify the regimes using expenditure effects. Intuitively, after identifying the second regime, we only have to determine which of the remaining sets is above the second regime, and which is below (see Figure 1.9). This result uses, but does not rely on, Assumption B.2(iv).

Identification of the Benefit Policy

Next, I show how to identify the policy $b(\cdot)$ in the second regime, and then use what we have learned from the second regime to identify bounds in the first and third regimes.

Lemma 1.3. Under Assumptions 1.1 to 1.7, N, and B.2:

$$y^*_\theta(w) = x^*_{\theta,2}(w) \;\text{ and }\; b^*_\theta(w) = px^*_{\theta,1}(w), \tag{1.3.6}$$

at each $w \in W^*_{\theta,2}$. Therefore, $y^*_\theta(\cdot)$ and $b^*_\theta(\cdot)$ are identified over $W^*_{\theta,2}$.

Proof. This result follows from the definition of the second regime in Corollary 1.2. □

Lemma 1.3 implies that, when demand is in the second regime, we can use what we observe to recover the household's income $y^*_\theta(w)$ and benefits $b^*_\theta(w)$. Lemma 1.3 follows from the fact that, in the second regime, the household consumes on the kink of its budget set. Next, consider the following implication:

Theorem 1.6. Under Assumptions 1.1 to 1.7, N, and B.2, the policy $b(\cdot)$ is identified over:

$$\big\{(y^*_\theta(w), p)' : w \in W^*_{\theta,2}\big\}. \tag{1.3.7}$$

Proof. By the definition of the pseudo-policy function $b^*_\theta(\cdot)$, we have: $b^*_\theta(w) = b(y^*_\theta(w), p)$. Moreover, by Proposition 1.8 and Lemma 1.3, we can invert the pseudo-income function $y^*_\theta(\cdot)$ to obtain the expenditure function $e_\theta(\cdot)$, at every z in the set in (1.3.7). Identification is, then, implied by the following relationship:

$$b(z) = b(y^*_\theta(e_\theta(z), p), p) = b^*_\theta(e_\theta(z), p). \tag{1.3.8}$$

□

Now, define:

$$e_\ell(p) \equiv \inf \left\{ e > 0 : (e, p) \in W^*_{\theta,2} \right\} \ \text{ and } \ e_h(p) \equiv \sup \left\{ e > 0 : (e, p) \in W^*_{\theta,2} \right\}. \tag{1.3.9}$$

Here, $e_\ell(p)$ is the *lowest* total expenditure e in the second regime given p, and $e_h(p)$ is the *highest* total expenditure e in the second regime given p. By Assumptions B.2(ii) and B.2(iv), and Theorem 1.5, these amounts are identified, for any observable p. By Lemma 1.3, we can also identify: $y^*_\theta(e_\ell(p), p)$, $b^*_\theta(e_\ell(p), p)$, $y^*_\theta(e_h(p), p)$, and $b^*_\theta(e_h(p), p)$.

Theorem 1.7. Under Assumptions 1.1 to 1.7, N, and B.2:

(i) $z \in R_{\theta,1}$ implies $\max\{0, e_h(p) - y\} \leq b(z) \leq b^*_\theta(e_h(p), p)$.

(ii) $z \in R_{\theta,3}$ implies $b^*_\theta(e_\ell(p), p) \leq b(z) < e_\ell(p) - y$.

Proof. See Appendix 1.B.11. □

From Theorem 1.7, we can identify bounds for the policy $b(\cdot)$ in both the first and third regimes, as long as the price p is observable (see Figure 1.10). Theorem 1.7 follows from Theorem 1.6 and the fact that the derivative of the policy $b(\cdot)$ is strictly larger than -1, and no larger than 0. Note, the first inequality in Theorem 1.7(i) is strict if $e_h(p) > y$.

Example 1.8 (Continued)**.** Suppose that the household has a Stone-Geary utility function, defined by: $u(x) = \sqrt{x_1 x_2}$, for every $x \in \bar{R}$. Further suppose that the policy $b(\cdot)$ has the form with fixed prices in (1.2.6). Under these specifications, the pseudo-policy function $b^*_\theta(\cdot)$ has the form in (1.2.39), and the pseudo-demand function $x^*_\theta(\cdot)$ has the form in (1.2.40). Thus, the definition of the second regime yields:

$$e_\ell(p) = \frac{(1+\pi)\gamma_1}{1+\pi\gamma_2} \ \text{ and } \ e_h(p) = \frac{2\gamma_1}{1+\gamma_2}, \tag{1.3.10}$$

for every $p > 0$. Consequently, we obtain:

$$b^*_\theta(e_\ell(p), p) = \frac{\gamma_1}{1+\pi\gamma_2} \ \text{ and } \ b^*_\theta(e_h(p), p) = \frac{\gamma_1}{1+\gamma_2}. \tag{1.3.11}$$

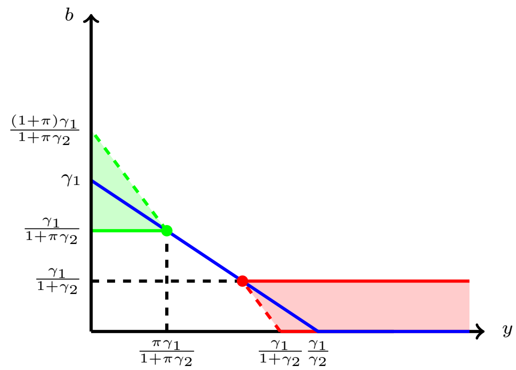

Figure 1.10. **Bounds for the Benefit Policy.** Bounds for the policy in (1.2.6) from Example 1.1. The blue line denotes the policy. The bounds for the policy in the first regime are in red. Likewise, the bounds for the policy in the third regime are in green.

Theorem 1.7 implies:

$$\max\left\{0, \frac{2\gamma_1}{1+\gamma_2} - y\right\} \leq b(z) \leq \frac{\gamma_1}{1+\gamma_2}, \tag{1.3.12}$$

for every $z \in R_{\theta,1}$. The lower bound is smaller than the upper bound in (1.3.12) because $\gamma_1 > 0$ and $z \in R_{\theta,1}$ if, and only if, $y > \frac{\gamma_1}{1+\gamma_2}$. In a similar fashion, Theorem 1.7 implies:

$$\frac{\gamma_1}{1+\pi\gamma_2} \leq b(z) \leq \frac{(1+\pi)\gamma_1}{1+\pi\gamma_2} - y, \tag{1.3.13}$$

for every $z \in R_{\theta,3}$. The lower bound is smaller than the upper bound in (1.3.13) because $z \in R_{\theta,3}$ if, and only if, $y < \frac{\pi\gamma_1}{1+\pi\gamma_2}$. I illustrate these bounds in Figure 1.10. In this figure, we can clearly see that the policy $b(\cdot)$ always lies between these bounds, and that, loosely speaking, these bounds become wider as y moves further from the boundary of the second regime. △

Identification of Standard Demand

We can also use what we know about the form of pseudo-demand $x^*_\theta(\cdot)$ to identify standard demand $x_u(\cdot)$ in the first regime. In particular, we obtain the following identification result:

Theorem 1.8. Under Assumptions 1.1 to 1.7, N, and B.2:

$$x^*_\theta(w) = x_u(w), \tag{1.3.14}$$

at each $w \in W^*_{\theta,1}$. Therefore, standard demand $x_u(\cdot)$ is identified over $W^*_{\theta,1}$.

Proof. By Corollary 1.2: $x^*_\theta(w) = x_u(\psi^*_\theta(w), p)$, for every $w \in W^*_{\theta,1}$. It is, therefore, left to show $e = \psi^*_\theta(w)$, for every $w \in W^*_{\theta,1}$. This result holds since:

$$e = px^*_{\theta,1}(w) + x^*_{\theta,1}(w) = px_{u,1}(\psi^*_\theta(w), p) + x_{u,2}(\psi^*_\theta(w), p) = \psi^*_\theta(w), \tag{1.3.15}$$

for every $w \in W^*_{\theta,1}$. □

It turns out, we can identify standard demand $x_u(\cdot)$ if, and only if, pseudo-demand $x^*_\theta(\cdot)$ is in the first regime, so that, when income y is not observed, we lose our ability to identify standard demand in the third regime. Hence, we can only deduce the exact change in demand from, say, a change in policy, if we are analyzing choices in the first regime. I omit the bounds for standard demand $x_u(\cdot)$ in the second and third regimes, for brevity.

Identification of the Discount Factor

We are now in a position to address the identification of the discount factor π. The following prerequisite result uses budget exhaustion and the definitions of $y^*_\theta(\cdot)$ and $b^*_\theta(\cdot)$:

Lemma 1.4. Under Assumptions 1.1 to 1.7, N, and B.2:

$$\frac{\partial x^*_{\theta,2}(w)}{\partial e} < \frac{\partial y^*_\theta(w)}{\partial e} < \frac{1}{1-\pi} \quad \text{and} \quad -\frac{1}{1-\pi} < \frac{\partial b^*_\theta(w)}{\partial e} \leq 0, \tag{1.3.16}$$

at each $w \in W^*_{\theta,3}$.

Proof. See Appendix 1.B.12. □

Lemma 1.4 implies that, in the third regime, (i) the pseudo-income function $y^*_\theta(\cdot)$ increases in e faster than pseudo-demand for non-food $x^*_{\theta,2}(\cdot)$, but slower than $\frac{1}{1-\pi}$, and (ii) the pseudo-policy function $b^*_\theta(\cdot)$ decreases in e at a rate no faster than $-\frac{1}{1-\pi}$. As in Theorem 1.7, we can integrate these bounds in order to construct bounds for the pseudo-income function $y^*_\theta(\cdot)$ and the pseudo-policy function $b^*_\theta(\cdot)$:

Lemma 1.5. Under Assumptions 1.1 to 1.7, N, and B.2:

$$\begin{aligned} &\max\left\{0, y^*_\theta(e_\ell(p),p) - \frac{|e_\ell(p)-e|}{1-\pi}\right\} < y^*_\theta(w) < x^*_{\theta,2}(w) \\ &\text{and} \;\; b^*_\theta(e_\ell(p),p) \leq b^*_\theta(w) < b^*_\theta(e_\ell(p),p) + \frac{|e_\ell(p)-e|}{1-\pi}, \end{aligned} \tag{1.3.17}$$

at each $w \in W^*_{\theta,3}$.

Proof. See Lemma 1.4. □

These bounds are not identified because they depend on the (unobserved) discount factor π. That being said, these bounds imply that, in the third regime, pseudo-income $y^*_\theta(w)$ is strictly larger than 0, and benefits $b^*_\theta(w)$ are no smaller than $b^*_\theta(e_\ell(p),p)$, which is identified. We can use this implication to bound the discount factor π above:

Theorem 1.9. Under Assumptions 1.1 to 1.7, N, and B.2:

$$\pi < \min\left\{1, \inf_{w\in W^*_{\theta,3}} \frac{x^*_{\theta,2}(w)}{b^*_\theta(e_\ell(p),p) - p x^*_{\theta,1}(w)}\right\} \equiv \pi^*_h. \tag{1.3.18}$$

Proof. See Appendix 1.B.13. □

Remark 1.5. Lemma 1.5 is the first noteworthy use of the fact that standard demand for non-food $x_{u,2}(\cdot)$ is strictly increasing in income y on $\mathbb{R}^2_{++}$. This assumption implies that pseudo-demand for non-food $x^*_{\theta,2}(\cdot)$ is strictly increasing in total expenditure e, a sufficient condition for the upper bound for the pseudo-income function $y^*_\theta(\cdot)$ in (1.3.17) to be the tightest upper bound that we can construct using Lemma 1.4 in the third regime. Without this assumption, Lemma 1.4 implies that $y^*_\theta(w)$ is only bounded above by:

$$y^*_\theta(e_\ell(p), p) - \int_e^{e_\ell(p)} \max\left\{0, \frac{\partial x^*_{\theta,2}(w)}{\partial e}\right\} de. \tag{1.3.19}$$

This bound is strictly smaller than the upper bound for $y^*_\theta(w)$ in (1.3.17) if, and only if, pseudo-demand for non-food $x^*_{\theta,2}(\cdot)$ is strictly increasing in total expenditure e on the interval $[e, e_\ell(p)]$.

Remark 1.6. The upper bound π^*_h for the discount factor π in (1.3.18) follows from the identifiable bounds for $y^*_\theta(\cdot)$ and $b^*_\theta(\cdot)$, as described in the discussion immediately following Lemma 1.5, and the fact that, by budget exhaustion:

$$\pi = \frac{x^*_{\theta,2}(w) - y^*_\theta(w)}{b^*_\theta(w) - px^*_{\theta,1}(w)}, \tag{1.3.20}$$

for every $w \in W^*_{\theta,3}$. Now, notice that, if the upper bound π^*_h for the discount factor π in (1.3.18) is strictly smaller than 1, then we can use Lemma 1.5 to construct a lower bound for pseudo-income $y^*_\theta(w)$ that is, at least for some pairs $w \in W_{\theta,3}$, strictly larger than 0 (i.e. the previous identifiable lower bound for pseudo-income $y^*_\theta(w)$) such that:

$$y^*_\ell(w) \equiv \max\left\{0, y^*_\theta(e_\ell(p), p) - \frac{|e_\ell(p) - e|}{1 - \pi^*_h}\right\} < y^*_\theta(w). \tag{1.3.21}$$

This new lower bound on pseudo-income $y^*_\theta(w)$ can, then, be used to construct a new upper bound for the discount factor π that is weakly smaller than the upper bound π^*_h:

$$\pi < \inf_{w \in W^*_{\theta,3}} \frac{x^*_{\theta,2}(w) - y^*_\ell(w)}{b^*_\theta(e_\ell(p), p) - px^*_{\theta,1}(w)}. \tag{1.3.22}$$

If this upper bound is strictly smaller than π^*_h, then we can repeat this procedure to construct a tighter lower bound for pseudo-income $y^*_\theta(w)$, and a tighter upper bound for the discount factor π. We can repeat this procedure until these bounds converge. I restrict attention to π^*_h, stopping at one interation of this procedure, for simplicity.

Example 1.8 (Continued)**.** Suppose that the household has a Stone-Geary utility function, defined by: $u(x) = \sqrt{x_1 x_2}$, for every $x \in \bar{R}$. Further suppose that the policy $b(\cdot)$ has the form with fixed prices in (1.2.6). Under these specifications, the amount $b^*_\theta(e_\ell(p), p)$ has the form in (1.3.11), and the pseudo-demand function $x^*_\theta(\cdot)$ has the form in (1.2.40). If we observe the entire third regime such that

$W^*_{\theta,3} = W_{\theta,3}$, then we obtain:

$$\pi^*_h = \min\left\{1, \inf_{w \in W_{\theta,3}} \frac{\pi e(1+\pi\gamma_2)}{\gamma_1(1+\pi) - e(1+\pi\gamma_2)}\right\}. \tag{1.3.23}$$

I obtain this expression by inputing $b^*_\theta(\cdot)$ and $x^*_\theta(\cdot)$ into (1.3.18), and then simplifying the resulting expression. Since the fraction in this expression is strictly increasing in e, it attains its infimum at the lowest admissible value of total expenditure $\frac{(1+\pi)\gamma_1}{2}$. Thus:

$$\pi^*_h = \min\left\{1, \frac{\pi(1+\pi\gamma_2)}{1-\pi\gamma_2}\right\}. \tag{1.3.24}$$

This bound is strictly smaller than 1 if, and only if:

$$\gamma_2 < \frac{1-\pi}{\pi(1+\pi)}. \tag{1.3.25}$$

In words, we can identify an "informative" upper bound π^*_h for the discount factor π if, and only if, the rate γ_2 at which the policy $b(\cdot)$ decreases with respect to income y is small, relative to a function of the discount factor π. Since the fraction on the right-hand side of this inequality is decreasing in π on $(0,1)$, this inequality is more likely to hold for smaller values of π. Intuitively, the smaller the discount factor π, the larger its effect on pseudo-demand $x^*_\theta(\cdot)$, improving our ability to make inference about π. △

Identification of Demand for Fraud

The upper bound π^*_h on the discount factor is needed to identify informative bounds for the demand for fraud $f_\theta(\cdot)$. For ease of exposition, I impose the following assumption:

Assumption 1.8. The upper bound π^*_h in (1.3.18) is strictly smaller than 1.

Assumption 1.8 implies that we can identify an informative upper bound for the discount factor π, as described in the previous section. As seen in Example 1.8, whether Assumption 1.8 holds depends on (i) the form of the benefit policy, (ii) the household's preferences, and (iii) what we observe. Fortunately, we can simply check Assumption 1.8 when π^*_h is identified, as under Assumptions 1.1 to 1.7, N, and B.2.

Corollary 1.4. Under Assumptions 1.1 to 1.8, N, and B.2:

$$\begin{gathered}\max\left\{0, y^*_\theta(e_\ell(p),p) - \frac{|e_\ell(p)-e|}{1-\pi^*_h}\right\} < y^*_\theta(w) < x^*_{\theta,2}(w) \\ \text{and } \; b^*_\theta(e_\ell(p),p) \le b^*_\theta(w) < b^*_\theta(e_\ell(p),p) + \frac{|e_\ell(p)-e|}{1-\pi^*_h},\end{gathered} \tag{1.3.26}$$

at each $w \in W^*_{\theta,3}$.

Proof. See Lemma 1.5 and Theorem 1.9. □

Corollary 1.4 follows from the fact that, if we replace the discount factor π in the bounds in (1.3.17) with the upper bound π_h^* in (1.3.18), then these bounds become wider. Since the upper bound π_h^* is identified, the bounds in Corollary 1.4 are also identified. We can, therefore, use these bounds to identify bounds for the pseudo-fraud function $f_\theta^*(\cdot)$:

Lemma 1.6. Under Assumptions 1.1 to 1.8, N, and B.2:

$$b_\theta^*(e_\ell(p), p) - px_{\theta,1}^*(w) \leq f_\theta^*(w) \leq b_\theta^*(e_\ell(p), p) + \frac{|e_\ell(p) - e|}{1 - \pi_h^*} - px_{\theta,1}^*(w). \tag{1.3.27}$$

for every $w \in W_{\theta,3}^*$.

Proof. See Corollaries 1.1 and 1.4. □

Lemma 1.6 implies that we can bound the household's chosen amount of fraud $f_\theta^*(w)$ both above and below, at every $w \in \mathcal{W}$. Under Assumptions 1.1 to 1.8, and N, these bounds are informative in the sense that, in the interior of the third regime, the lower bound is larger than 0, and the upper bound is finite.

While Lemma 1.6 is useful for the econometrician, it cannot directly be used to map designs $z \in \mathbb{R}_{++}^2$ to fraud $f_\theta(z)$, making it less useful for the FNS. That being said, Lemma 1.6 provides the foundation for the identification of bounds for the demand for fraud $f_\theta(\cdot)$. First, we need to identify bounds for total expenditure $e_\theta(\cdot)$ as a function of $z = (y, p)$.

Lemma 1.7. Under Assumptions 1.1 to 1.8, N, and B.2:

$$\lambda_\ell(z) \equiv \{e : x_{\theta,2}^*(w) = y\} < e_\theta(z) < (1 - \pi_h^*)\big[y - y_\theta^*(e_\ell(p), p)\big] + e_\ell(p) \equiv \lambda_h(z), \tag{1.3.28}$$

for every $z \in R_{\theta,3}$.

Proof. This result follows from rearranging the bounds for $y_\theta^*(z)$ in Corollary 1.4. □

Lemma 1.7 says that, in the third regime, total expenditure $e_\theta(z)$ is both (i) bounded below by the inverse $\lambda_\ell(z)$ of pseudo-demand for non-food $x_{\theta,2}^*(\cdot)$, given p, evaluated at y, and (ii) bounded above by a linear transformation $\lambda_h(z)$ of the distance between y and $y_\theta^*(e_\ell(p), p)$ that depends on the bound π_h^*.

Theorem 1.10. Under Assumptions 1.1 to 1.8, N, and B.2:

$$\begin{aligned} b_\theta^*(e_\ell(p), p) - px_{\theta,1}^*(\lambda_h(z), p) &\leq f_\theta(z) \\ \leq b_\theta^*(e_\ell(p), p) + \frac{|e_\ell(p) - \lambda_\ell(z)|}{1 - \pi_h^*} &- px_{\theta,1}^*(\lambda_\ell(z), p), \end{aligned} \tag{1.3.29}$$

for every $z \in R_{\theta,3}$.

Proof. See Lemma 1.6 and 1.7. □

Theorem 1.10 implies that we can bound the household's demand for fraud $f_\theta(z)$ both above and below, at every $z \in R_{\theta,3}$ at which $(\lambda_\ell(z), p)', (\lambda_h(z), p)' \in \mathcal{W}$. Without Assumption 1.8, these bounds become uninformative—in particular, the lower bound becomes 0, and the upper bound becomes infinite.

Remark 1.7. In practice, we also observe the maximum possible amount of benefits (see Table 1.11 in Appendix 1.A). We could, theoretically, apply this information to construct upper bounds on, say, the pseudo-fraud function $f^*_\theta(\cdot)$ if Assumption 1.8 is violated. For instance, if $b_h(p)$ denotes this maximum, then, for every $w \in W^*_{\theta,3}$, we must have:

$$f^*_\theta(w) \le b_h(p) - px^*_{\theta,1}(w). \tag{1.3.30}$$

Example 1.8 (Continued)**.** Suppose that the household has a Stone-Geary utility function, defined by: $u(x) = \sqrt{x_1 x_2}$, for every $x \in \bar{R}$. Further suppose that the policy $b(\cdot)$ has the form with fixed prices in (1.2.6). Under these specifications, the pseudo-income function $y^*_\theta(\cdot)$ has the form in (1.2.35) and the pseudo-policy function $b^*_\theta(\cdot)$ has the form in (1.2.39). Furthermore, the pseudo-demand function $x^*_\theta(\cdot)$ has the form in (1.2.40), and we obtain the amounts in (1.3.10) and (1.3.11), and the bound in (1.3.24). Moreover, we obtain:

$$y^*_\theta(e_\ell(p), p) = \frac{\pi\gamma_1}{1+\pi\gamma_2} \quad \text{and} \quad y^*_\theta(e_h(p), p) = \frac{\gamma_1}{1+\gamma_2}, \tag{1.3.31}$$

and the expenditure function $e_\theta(\cdot)$ is bounded by:

$$\lambda_\ell(z) = \frac{(1+\pi)y}{\pi} \quad \text{and} \quad \lambda_h(z) = \frac{(1+\pi)\gamma_1 - (1-\pi^*_h)[\pi\gamma_1 - y(1+\pi\gamma_2)]}{1+\pi\gamma_2}. \tag{1.3.32}$$

in the third regime. Therefore:

$$\frac{(1-\pi^*_h)[\pi\gamma_1 - (1+\pi\gamma_2)y]}{(1+\pi)(1+\pi\gamma_2)} \le f_\theta(z) \le \frac{(2-\pi-\pi^*_h)[\pi\gamma_1 - (1+\pi\gamma_2)y]}{\pi(1-\pi^*_h)(1+\pi\gamma_2)}, \tag{1.3.33}$$

for every $z \in R_{\theta,3}$. It is easy to show that (i) these bounds are strictly positive in the interior of the third regime, and (ii) these bounds contain the demand for fraud $f_\theta(z)$ in (1.2.30) whenever the condition ensuring that π^*_h is smaller than 1 in (1.3.25) holds. △

1.3.4 Identification of Utility

In Sections 1.3.2 and 1.3.3, I explicitly focused on the non-parametric identification of specific latent functional parameters in the model of benefit fraud in Section 1.2, and some functions of these parameters. Here, I focus on the partial non-parametric identification of a determinant of demand: utility.

The objective of this section is somewhat analogous to the objective of identification in a *deep learning model*. Loosely speaking, deep learning models use observable data to identify components in a first layer of the model, and then use these objects to identify objects in a second layer, and continue this procedure until all layers have been identified. In this section, I use the fact that standard demand $x_u(\cdot)$ is identified to identify utility $u(\cdot)$.

The identification of utility $u(\cdot)$ involves solving an ordinary differential equation. The procedure in this section is closely related to the *integrability theorem* (see, for example, pages 243-245 in Samuelson, 1948, Theorem 2 in Hurwicz and Uzawa, 1971, Theorem 2 in Hosoya, 2013, and Section 2.4 in Hosoya, 2016), as well as *recoverability* (Mas-Colell, 1977).

An Ordinary Differential Equation

Recall, under Assumption 1.6, the subset $G(v)$ is a twice-continuousously-differentiable indifference curve, as long as it is non-empty. For exposition, let $g(\cdot, v)$ denote the indifference curve $G(v)$, written as a function of the quantity of food x_1 such that:

$$u(x_1, g(x_1, v)) = v, \tag{1.3.34}$$

for every $x_1 > 0$ and $v \neq u(0,0)$. By differentiating both sides of this equality with respect to the quantity of food x_1, we obtain:

$$\frac{\partial g(x_1, v)}{\partial x_1} = -m(x_1, g(x_1, v)), \tag{1.3.35}$$

where $m(x) \equiv \frac{\partial u(x)/\partial x_1}{\partial u(x)/\partial x_2}$ denotes the household's *marginal rate of substitution* evaluated at bundle $x \in R$: the rate at which the household is willing to exchange food for non-food given x. If we fix $g(x_1', v) = x_2'$, for some $x' \in R$, then (1.3.35) becomes an initial value problem. If this initial value problem has a unique global solution, then the solution coincides with the indifference curve that passes through x'. It is left to show that (i) we can identify the marginal rate of substitution $m(\cdot)$, and (ii) after we fix $g(x_1', v) = x_2'$, the ordinary differential equation in (1.3.35) has a unique global solution.

The Marginal Rate of Substitution

Consider the standard utility maximization problem under a linear budget constraint in (1.2.27). Under Assumption 1.6, standard demand $x_u(z)$ solves the following system of equations:

$$m(x) = p \;\text{ and }\; px_1 + x_2 = y. \tag{1.3.36}$$

The first equality implies that the slope of the indifference curve that passes through standard demand $x_u(z)$ equals the slope of the boundary of the household's budget set, also known as the budget line; the second equality is Walras' law (Walras, 1874). The first equality implies that, if we can invert standard demand $x_u(\cdot)$, then the second component of this inverse:

$$z_u(x) = \begin{bmatrix} y_u(x) \\ p_u(x) \end{bmatrix}, \tag{1.3.37}$$

which characterizes the price p as a function of x, coincides with the marginal rate of substitution such that $m(x) = p_u(x)$. Since the invertibility of standard demand $x_u(\cdot)$ follows from strong quasi-concavity (see Proposition 2 in Chapter 2), if standard demand $x_u(\cdot)$ is identified over a subset $\mathcal{Z}_0$, then the marginal rate of substitution $m(\cdot)$ is identified over the range of standard demand $x_u(\cdot)$ on this subset.

Existence of a Unique Global Solution

Under Assumption 1.6, the marginal rate of substitution $m(\cdot)$ is continuously-differentiable on R. As a result, the Picard-Lindelöf theorem implies that the ordinary differential equation in (1.3.35) has a unique local solution, for every initial condition. It is, therefore, left to find conditions on the range of standard demand $x_u(\cdot)$ over its identified set $\mathcal{Z}_0$ under which this solution can be analytically extended to the boundary of this range in a unique way. It is sufficient for this range to be open and for the intersection of this range with the indifference curve $G(v)$ to be a connected set (see Section 2.3 in Chapter 2 for more discussion, and Section 1.3.4 for a discussion of what happens when this intersection is not connected.

Remark 1.8. Here, the existence of a unique global solution is simplified by the fact that there are only two distinct goods. When there are three or more distinct goods, we require an "integrability condition" (see Theorem 10.9.4 in Dieudonné, 1960, for a theoretical result, and Section 2 in Samuelson, 1950, for a discussion of this feature).

Main Result

Sections 1.3.4 to 1.3.4 describe a procedure that can be used to identify utility $u(\cdot)$ from standard demand $x_u(\cdot)$. I now formally summarize this result. Consider an assumption:

Assumption B.3.

(i) Standard demand $x_u(\cdot)$ is identified over $\mathcal{Z}_0$.

(ii) The closure $\mathcal{X}_0$ of the range $x_u(\mathcal{Z}_0)$ admits an open subset that is dense in $\mathcal{X}_0$.

(iii) The following set is connected:

$$\mathcal{X}_{1,v} = \big\{x_1 \geq 0 : (x_1, x_2) \in \mathcal{X}_0 \cap G(v), \text{ for some } x_2 \geq 0\big\}. \tag{1.3.38}$$

Assumption B.3(ii) implies that the closure of the interior of $x_u(\mathcal{Z}_0)$ equals the closure of $x_u(\mathcal{Z}_0)$. Assumption B.3(ii) is slightly more general than assuming that $x_u(\mathcal{Z}_0)$ is open. I do not assume that

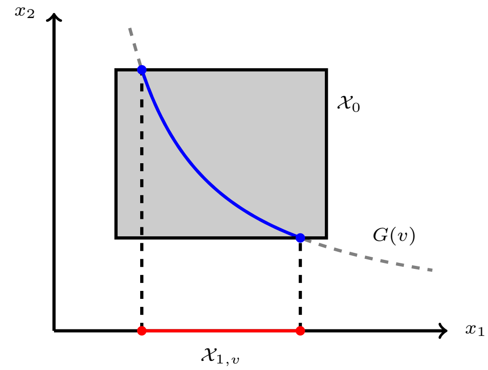

Figure 1.11. **Sets in Assumption B.3.** The shaded region is $\mathcal{X}_0$. The blue curve is the intersection of $\mathcal{X}_0$ and $G(v)$. The red line is $\mathcal{X}_{1,v}$.

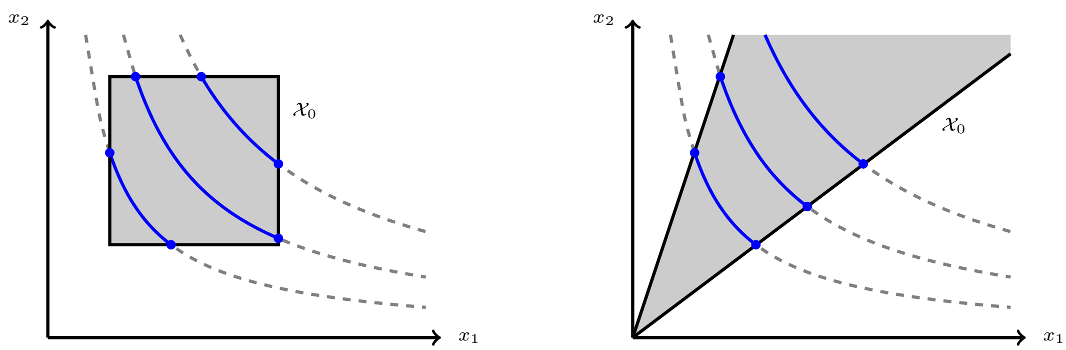

Figure 1.12. **Connected Sets.** In each figure, the shaded region is $\mathcal{X}_0$ and the blue curves denote the intersection of $\mathcal{X}_0$ with an indifference curve. These intersections are connected whenever $\mathcal{X}_0$ is a rectangle, or a cone, because indifference curves are downward sloping and strictly convex.

$x_u(\mathcal{Z}_0)$ is open because, in general, it will not be open, even if $\mathcal{Z}_0$ is open. However, under Assumption B.3(ii), if we can identify utility $u(\cdot)$ over the interior of $x_u(\mathcal{Z}_0)$, then we can extend this identification to $\mathcal{X}_0$ using the continuity of utility $u(\cdot)$, as described in Remark 1.4. Assumption B.3(iii) says that the "observable" part of the indifference curve $G(v)$ is connected—the set $\mathcal{X}_{1,v}$ is the projection of the intersection of $\mathcal{X}_0$ and $G(v)$ onto the x_1-axis. Assumption B.3(iii) ensures that $g(\cdot, v)$ never leaves and re-enters $\mathcal{X}_0$. Many conditions can guarantee that this property holds. For example, it is sufficient for $\mathcal{Z}_0$ to be a rectangle, or a cone, in $\mathbb{R}^2_{++}$ (see Figure 1.12).

Theorem 1.11. Under Assumptions 1.1, 1.6, and B.3:

(i) The marginal rate of substitution $m(\cdot)$ is identified over $\mathcal{X}_0$.

(ii) The indifference curve $g(\cdot, v)$ is identified over $\mathcal{X}_{1,v}$.

Proof. See Theorem 1 in Chapter 2. □

Theorem 1.11(i) does not use the restrictions in Assumption B.3(ii). Theorem 1.11(ii) says that $G(v)$ is identified where it intersects $\mathcal{X}_0$. Theorem 1.11 says that utility $u(\cdot)$ is identified up to a strictly increasing transformation over $\mathcal{X}_0$. While Theorem 1.11(ii) is related to the integrability theorem, I assume that standard demand $x_u(\cdot)$ is generated by a well-behaved utility function, rather than proving that such a function exists. Theorem 1.11 should also be distinguished from finite sample methods (see Afriat, 1967, and Varian, 1982). To summarize, the steps:

(i) Invert standard demand $x_u(\cdot)$ to recover $m(x) = p_u(x)$ over $\mathcal{X}_0$.

(ii) Fix $x' \in \mathcal{X}_0$ and solve:

$$\frac{\partial g(x_1, v)}{\partial x_1} = -m(x_1, g(x_1, v)) \quad \text{subject to} \quad g(x_1', v) = x_2', \tag{1.3.39}$$

for every $x_1 \in \mathcal{X}_{1,v}$. The solution extends to the boundary of the observable set $\mathcal{X}_{1,v}$ and it coincides with the indifference curve $g(\cdot, v)$ that passes through $x' \in \mathcal{X}_0$.

Example 1.8 (Continued)**.** Suppose that the household has a Stone-Geary utility function, defined by: $u(x) = \sqrt{x_1 x_2}$, for every $x \in \bar{R}$. Under this specification, the marginal rate of substitution has the form: $m(x) = x_2/x_1$, for each $x \in R$. Moreover, standard demand has the form: $x_u(z) = \left(\frac{y}{2p}, \frac{y}{p}\right)'$, for each $z \in \mathbb{R}^2_{++}$. Inverting yields:

$$z(x) = \left(2x_2, \frac{x_2}{x_1}\right)', \tag{1.3.40}$$

As expected, the second component of this inverse coincides with the marginal rate of substitution $m(\cdot)$. Now, notice that, if standard demand $x_u(\cdot)$ is identified over the entire orthant R, then $\mathcal{X}_{1,v}$ coincides with the set of positive real-numbers $\mathbb{R}_{++}$. For this utility function, the ordinary differential equation in (1.3.35) becomes a linear differential equation such that:

$$\frac{\partial g(x_1, v)}{\partial x_1} = -\frac{g(x_1, v)}{x_1}. \tag{1.3.41}$$

The solution to this differential equation has the form: δ_v/x_1, in which δ_v denotes a function of the integrating constant. The form of this solution is found by dividing (1.3.41) by $g(x_1, v)$, integrating, and appling the exponential transform. Consequently, if we solve (1.3.41) subject to $g(x_1', v) = x_2'$, for some $x' \in R$, then $\delta_v = x_1' x_2'$. To see that this solution equals the indifference curve that passes through x', notice that, under this specification (i) $g(x_1, v) = v^2/x_1$, and (ii) $v = u(x')$ implies $v^2 = x_1' x_2'$. △

Discussion

Theorem 1.11 implies that utility $u(\cdot)$ is identified up to a strictly increasing transformation over $\mathcal{X}_0$, but it says nothing about what happens (i) outside the range $\mathcal{X}_0$, or (ii) when $\mathcal{X}_{1,v}$ is not connected. In each case, we can find a partial order. For example, in Figure 1.13, we can use the fact that the indifference curves of $u(\cdot)$ are downward sloping and strictly convex to infer that the unobserved continuation of $G(v_0)$ lies in the red and orange regions. Similarly, we can use the monotonicity and transitivity of

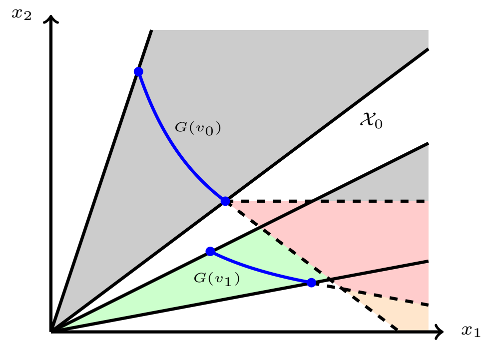

Figure 1.13. **Identification Outside $\mathcal{X}_0$.** The gray regions are $\mathcal{X}_0$. The blue curves denote intersections of $\mathcal{X}_0$ with indifference curves. The continuation of $G(v_0)$ outside $\mathcal{X}_0$ lies in the red and orange regions because the indifference curves of $u(\cdot)$ are downward sloping and strictly convex. By monotonicity and transitivity, every bundle on $G(v_0)$ is strictly better than every bundle in the green region. We can then use this information to infer that the continuation of $G(v_0)$ outside $\mathcal{X}_0$ lies in the red region.

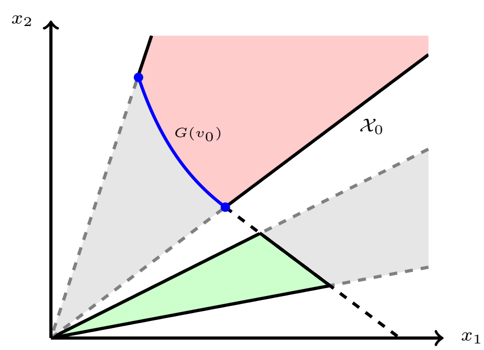

Figure 1.14. **Recovering a Total Order on Disjoint Sets.** The union of red and green regions is $\mathcal{X}_0$. The blue curve denotes the intersection of $\mathcal{X}_0$ with an indifference curve. This indifference curve chracterizes the set of least preferred bundles in the red region. We can recover a total order because the least preferred bundles in the red region are preferred to all bundles in the green region.

utility $u(\cdot)$ to infer that every bundle on $G(v_0)$ is strictly better than every bundle in the green region, including those along $G(v_1)$. We can, then, use this information to tighten the lower bound on $G(v_0)$ outside $\mathcal{X}_0$. This procedure implies that the unobserved continuation of $G(v_0)$ lies in the red region. Since this argument holds for every indifference curve that intersects the green region, we can choose an indifference curve that yields a tight lower bound. We can also use a similar argument for the upper bound, and continue this procedure until we obtain sharp bounds. This argument also implies that there are some exceptions in which we can recover a total order over disjoint sets—for example, in Figure 1.14, the least preferred bundle in the red region is preferred to every bundle in the green region, implying that we can recover a total order. While I do not discuss these aspects further, they are related to revealed preference, and what arises with finitely many observations of standard demand $x_u(\cdot)$ (see Afriat, 1967, or Varian, 1982).

Link with Other Identification Results

Let us now consider the link between the identification result in this section and the results from Sections 1.3.2 and 1.3.3—that is, let us discuss the appropriateness of the assumptions on the observable set $\mathcal{X}_0$.

If income y is observed, we can identify standard demand $x_u(\cdot)$ over:

$$\mathcal{Z}_0^* \equiv \left\{(\psi(z), p)' : z \in R_{\theta,1}^*\right\} \cup \left\{(\phi(z), \pi p)' : z \in R_{\theta,3}^*\right\}. \tag{1.3.42}$$

Theorem 1.11 implies that we can identify utility $u(\cdot)$ over the closure $\mathcal{X}_0^*$ of the range $x_u(\mathcal{Z}_0^*)$ if (i) $x_u(\mathcal{Z}_0^*)$ admits an open subset that is dense in $\mathcal{X}_0^*$, and (ii) the following set is connected:

$$\mathcal{X}_{1,v}^* \equiv \left\{x_1 \geq 0 : (x_1, x_2) \in \mathcal{X}_0^* \cap G(v), \text{ for some } x_2 \geq 0\right\}, \tag{1.3.43}$$

for every observable (indirect) utility level $v \neq u(0,0)$. While the first condition holds under weak assumptions (see Remark 1.4), the second condition does not usually hold—in general, the projection $\mathcal{X}_{1,v}^*$ is disconnected. However, even if $\mathcal{X}_{1,v}^*$ is disconnected, we can identify utility $u(\cdot)$ over a proper subset of $\mathcal{X}_0^*$ on which these conditions hold (and deduce a partial order over $\mathcal{X}_0^*$, if desired, as in Section 1.3.4).

The second condition—that is, the connectedness of the projection of the identified range of standard demand—is less worrisome when income y is not observed because we can only identify standard demand in the first regime. Indeed, the second condition is usually violated when income y is observed because $\mathcal{Z}_0^*$ is the union of two disjoint sets. When income y is not observed, we only need to ensure that the set associated with the first regime $W_{\theta,1}^*$ is well-behaved (recall the conversation after Assumption B.3).

Example 1.8 (Continued)**.** Suppose that the household has a Stone-Geary utility function, defined by: $u(x) = \sqrt{x_1 x_2}$, for every $x \in \bar{R}$. Under this specification, standard demand has the form: $x_u(z) = \left(\frac{y}{2p}, \frac{y}{2}\right)'$, for each $z \in \mathbb{R}_{++}^2$. Further suppose that the policy $b(\cdot)$ has the form with fixed prices in (1.2.6) subject to $\gamma_2 < 1$. These assumptions imply that the pseudo-demand function $x_\theta^*(\cdot)$ has the form in (1.2.40). Theorem 1.4 implies that, if we observe demand $x_\theta(\cdot)$ on R, then standard demand $x_u(\cdot)$ is identified over:

$$\mathcal{Z}_0^* = \left\{z \in \mathbb{R}_{++}^2 : y \in \left[\pi\gamma_1, \frac{2\pi\gamma_1}{1+\pi\gamma_2}\right] \cup \left[\frac{2\gamma_1}{1+\gamma_2}, \infty\right)\right\}. \tag{1.3.44}$$

Therefore, $\mathcal{X}_0^*$ is the Cartesian product of $[0,\infty)$ and:

$$\left[\frac{\pi\gamma_1}{2}, \frac{\pi\gamma_1}{1+\pi\gamma_2}\right) \cup \left[\frac{\gamma_1}{1+\gamma_2}, \infty\right). \tag{1.3.45}$$

In words, the closure of the identified range of standard demand $x_u(\cdot)$ over $\mathcal{Z}_0^*$ consists of all bundles $x \in \bar{R}$ for which x_2 is in the set in (1.3.45). We cannot identify utility $u(\cdot)$ over this set since it is disconnected, and it does not satisfy the exception in Section 1.3.4.

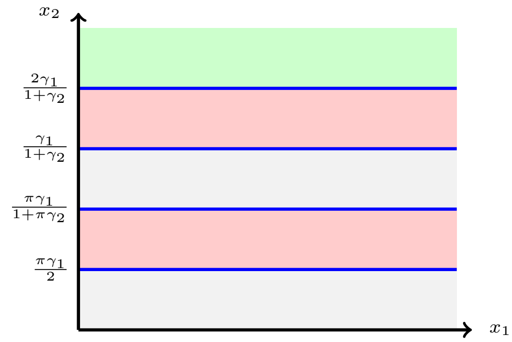

Figure 1.15. **Link with Other Identification Results.** This figure illustrates the sets from Example 1.8. When income y is observed, the identified range of standard demand $x_u(\cdot)$ in (1.3.45) is the union of the red and green regions. When income y is not observed, the identified range of standard demand $x_u(\cdot)$ in (1.3.47) is just the green region.

Now, suppose that we observe pseudo-demand $x_\theta^*(\cdot)$ at every admissible $w \in \mathbb{R}^2_{++}$. Then, by Theorem 1.8, standard demand $x_u(\cdot)$ is identified over:

$$W_{\theta,1}^* = \left\{ w \in \mathbb{R}^2_{++} : e > \frac{2\gamma_1}{1+\gamma_2} \right\}. \tag{1.3.46}$$

Therefore, $\mathcal{X}_0^*$ is the Cartesian product of $[0, \infty)$ and:

$$\left[\frac{\gamma_1}{1+\gamma_2}, \infty \right). \tag{1.3.47}$$

Unlike before, this range is an open rectangle in $\mathbb{R}^2_{++}$. Thus, utility $u(\cdot)$ is identified over the closure of the range of standard demand $x_u(\cdot)$ over $W_{\theta,1}^*$. I display the sets in (1.3.45) and (1.3.47) in Figure 1.15. $\triangle$

1.4 Statistical Inference

Let us now consider the non-parametric estimation of the identifiable functional parameters in a stochastic environment. I assume that we observe panel data, indexed by households i and months t. In the application in Section 1.5, I observe consumption choices x_{it}, expenditures e_{it}, and prices p_{it}, for a large number n of households, and a small number T of months. I, therefore, state results in terms of pseudo-demand $x_\theta^*(\cdot)$ (treating demand as a function of expenditure e, rather than income y), and consider n tending to infinity with a fixed number of months T.

The organization of the remainder of this section is as follows: First, I show how to derive observed prices and quantities from detailed scanner data. Formally, I show how to aggregate prices and quantities into two groups—namely, food and non-food—using the Laspeyres and Paasche indices. Next, I introduce the stochastic assumptions on the observations and latent stochastic model. Third, I introduce a non-parametric spline estimator for the pseudo-demand function $x_\theta^*(\cdot)$. Last, I discuss how to use the identification strategy in Section 1.3 to estimate some latent functional parameters.

1.4.1 Prices and Quantities

In practice, we do not directly observe the normalized price of food p_{it}, or the purchased quantities of food x_{i1t} and non-food x_{i2t}. Instead, these goods are comprised of many homogeneous goods. For instance, the dataset used in Section 1.5 contains information on "three million unique [universal product codes] for 1073 products in 106 product groups" (see Ng, 2017, and Guha and Ng, 2019). For each household i and month t, we observe a price p_{ijkt} and quantity x_{ijkt}, for every $k = 1, \ldots, K_j$ and $j = 1, 2$. To be precise, p_{i1kt} denotes the price of the k^{th} good classified as food, faced by household i in month t, and p_{i2kt} denotes the price of the k^{th} good classified as non-food, faced by household i in month t. The indices on x_{ijkt} have similar interpretations. Here, I show how to transform these observations with a common quantity unit, and, then, define aggregate prices p_{ijt} and quantities x_{ijt}. From this transformed data, we can construct the normalized price of food $p_{it} = p_{i1t}/p_{i2t}$ and expenditure $e_{it} = p_{it}x_{i1t} + x_{i2t}$.

When transforming the data, it is extremely important to avoid artificially forcing prices and/or expenditures to be the same across households (see Blundell et al., 2003, and Kitamura and Stoye, 2018, for examples of this practice). Indeed, prices vary significantly across retailers (even for identical goods), expenditures vary significantly across households, and the identification strategy described in Section 1.3 relies heavily on the existence of variation in expenditures.

To aggregate goods, I make use of the Laspeyres and Paasche indices. Both indices measure prices relative to a *benchmark*. Formally, let us consider average expenditure across all households in aggregate good group j in month 0:

$$E_{j0} \equiv \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K_j} p_{ijk0} x_{ijk0}. \tag{1.4.1}$$

This expression can be rewritten:

$$E_{j0} = \sum_{k=1}^{K_j} \left\{ \left(\frac{1}{n} \sum_{i=1}^{n} p_{ijk0} \right) \left(\sum_{i=1}^{n} p_{ijk0} x_{ijk0} \right) \left(\sum_{i=1}^{n} p_{ijk0} \right)^{-1} \right\} = \sum_{k=1}^{K_j} p_{jk0} x_{jk0}, \tag{1.4.2}$$

where p_{jk0} denotes the average price of the k^{th} good in aggregate group j in month 0:

$$p_{jk0} = \frac{1}{n} \sum_{i=1}^{n} p_{ijk0}, \tag{1.4.3}$$

and x_{jk0} denotes a benchmark bundle:

$$x_{jk0} = \left(\sum_{i=1}^{n} p_{ijk0} x_{ijk0} \right) \left(\sum_{i=1}^{n} p_{ijk0} \right)^{-1}. \tag{1.4.4}$$

Now, consider the ratio of household i's expenditure $e_{ijt} = \sum_{k=1}^{K_j} p_{ijkt}x_{ijkt}$ in aggregate group j in month t, and benchmark expenditure across all households E_{j0} from (1.4.1):

$$\frac{e_{ijt}}{E_{j0}} = \left(\sum_{k=1}^{K_j} p_{ijkt}x_{ijkt}\right)\left(\sum_{k=1}^{K_j} p_{jk0}x_{jk0}\right)^{-1}. \tag{1.4.5}$$

By definition, this expression is the product of the Laspeyres (price) index $\mathcal{L}_{ijt}$ and the Paasche (quantity) index $\mathcal{P}_{ijt}$ such that:

$$\frac{e_{ijt}}{E_{j0}} = \mathcal{L}_{ijt}\mathcal{P}_{ijt}, \tag{1.4.6}$$

where these indices are defined by:

$$\mathcal{L}_{ijt} = \frac{1}{E_{j0}}\left(\sum_{k=1}^{K_j} p_{ijkt}x_{jk0}\right) \text{ and } \mathcal{P}_{ijt} = e_{ijt}\left(\sum_{k=1}^{K_j} p_{ijkt}x_{jk0}\right)^{-1}. \tag{1.4.7}$$

Loosely speaking, the Laspeyres index $\mathcal{L}_{ijt}$ characterizes the relative evolution of the aggregate price between the pair (p_{ijkt}, x_{ijkt}) and our "benchmark." The interpretation of the Paasche index $\mathcal{P}_{ijt}$ is similar, but it characterizes the evolution of aggregate quantities. To illustrate, suppose $\mathcal{L}_{ijt} = 2$. This value for $\mathcal{L}_{ijt}$ implies that the benchmark bundle x_{jk0} costs twice as much in month t than in month 0. Similarly, $\mathcal{P}_{ijt} = 2$ implies that the bundle x_{ijkt} from month t costs twice as much as the benchmark bundle x_{ij0} in month t. Because these indices characterize the relative evolutions of prices and quantities, we expect to have aggregate prices and quantities that satisfy:

$$p_{ijt} = P_j\mathcal{L}_{ijt} \text{ and } x_{ijt} = X_j\mathcal{P}_{ijt}, \tag{1.4.8}$$

in which P_j and X_j denote values to be fixed, for each aggregate group $j = 1, 2$. For p_{ijt} and x_{ijt} to be reasonable measures for the price and quantity of aggregate good j in month t, we also need P_j and X_j to satisfy: $E_{j0} = P_jX_j$, for each aggregate group $j = 1, 2$. This restriction follows from the coherency of the definition of expenditure in aggregate group j—to be more precise, we expect the following amounts to coincide:

$$e_{ijt} = E_{j0}\mathcal{L}_{ijt}\mathcal{P}_{ijt} \text{ and } p_{ijt}x_{ijt} = P_jX_j\mathcal{L}_{ijt}\mathcal{P}_{ijt}. \tag{1.4.9}$$

That being said, even if we impose these equalities, there exists a remaining degree of freedom. Indeed, the goods that make up an aggregate group are almost always measured in qualitatively different units—for example, bottles, kilograms, or packs of six. We need to account for this degree of freedom by defining an Artificial Quantity Unit (AQU). Without loss of generality, we can choose an AQU by fixing $X_j = 1$, for each aggregate group $j = 1, 2$. Together, these two restrictions yield:

$$p_{ijt} = \sum_{k=1}^{K_j} p_{ijkt} x_{jk0} \;\text{ and }\; x_{ijt} = e_{ijt} \left(\sum_{k=1}^{K_j} p_{ijkt} x_{jk0} \right)^{-1}. \tag{1.4.10}$$

Therefore, the aggregate price p_{ijt} is the cost of the benchmark bundle x_{jk0} in month t, and the aggregate quantity x_{ijt} is expenditure e_{ijt}, relative to the cost of the benchmark bundle x_{jk0} in month t.

Remark 1.9. *I have used the Laspeyres-Paasche decomposition in a non-standard way.* This decomposition is usually used to aggregate across both goods and households (to measure inflation or deflation). Here, I do not aggregate across households to avoid artificially forcing prices and/or expenditures to be the same across households.

Remark 1.10. In general, it is not appropriate to aggregate by averaging prices and summing quantities within aggregate groups. Formally, this type of aggregation would yield expenditures $\bar{e}_{ijt}$ with the form:

$$\bar{e}_{ijt} = \frac{1}{K_j} \left(\sum_{k=1}^{K_j} p_{ijkt} \right) \left(\sum_{k=1}^{K_j} x_{ijkt} \right) = \frac{1}{K_j} \sum_{k=1}^{K_j} \sum_{\ell=1}^{K_j} p_{ijkt} x_{ij\ell t}. \tag{1.4.11}$$

These expenditures are rarely equal to e_{ijt}, and can even move in a different direction than e_{ijt}. For example, if aggregate good group j is comprised of $K_j = 2$ goods, and:

$$p_{ij11} = p_{ij21} = x_{ij11} = x_{ij21} = 1, \;\; p_{ij12} = x_{ij22} = 2, \;\text{ and }\; p_{ij22} = x_{ij12} = \frac{1}{2}, \tag{1.4.12}$$

then we obtain constant expenditure with $e_{ij1} = e_{ij2} = 2$, but a strict increase in $\bar{e}_{ijt}$:

$$\bar{e}_{ij1} = \frac{1}{2}(1+1)(1+1) = 2 < \frac{25}{8} = \frac{1}{2}\left(2 + \frac{1}{2}\right)\left(\frac{1}{2} + 2\right) = \bar{e}_{ij2}. \tag{1.4.13}$$

1.4.2 Latent Stochastic Model

Now that we have defined coherent notions of e_{it}, p_{it}, and x_{it}, we can formally describe the assumptions on the observations and latent stochastic model.

Assumption B.4. We jointly observe $w_{it} = (e_{it}, p_{it})$ and x_{it}, constrained by budget exhaustion $p_{it}x_{i1t} + x_{i2t} = e_{it}$, for every household $i = 1, \ldots, n$ and month $t = 1, \ldots, T$.

Assumption B.4 implicitly requires every household to consume positive quantities of the goods used to construct the normalized price of food p_{it} using the method described in Section 1.4.1, in each month (see Section IV.A in Blundell, Horowitz, and Parey, 2017, and Section 2.4.1 in Chapter 2, for two examples of this type of restriction in non-parametric settings). Because these goods are aggregate, it is assumed that we can drop the households violating this requirement without inducing a selection bias.

Let $Q_{x_{i1t}}(\tau|w_{it})$ denote the τ^{th} quantile of consumption x_{i1t} conditional on w_{it} such that $0 < \tau < 1$.[9]

Assumption 1.9.

(i) Pairs (w_{it}, x_{it}) are independently and identically distributed.

(ii) There exists $\tau \in (0,1)$ such that $Q_{x_{i1t}}(\tau|w_{it}) = x^*_{\theta,1}(w_{it})$, for some pseudo-demand function $x^*_\theta(\cdot)$.

Assumption 1.9 says that the τ^{th} conditional quantile of consumption is a pseudo-demand function. We can focus on the consumption of food x_{i1t} because there exists a deterministic one-to-one relationship between the components of consumption x_{it} given w_{it}. I focus on the τ^{th} conditional quantile, without providing a specific numerical value for τ (by, for example, restricting attention to the median) in order to maintain generality, and facilitate some discussions that follow. Assumption 1.9 does not assume that *every* conditional quantile of consumption $Q_{x_{i1t}}(\tau|w_{it})$ yields a pseudo-demand function.

It is important to notice that, Assumption 1.9 does not assume that the *conditional mean* of consumption is a pseudo-demand function (with the form in Section 1.2.12). Indeed, it would be unreasonable to impose such an assumption. For example, if preferences are stochastic, independent of pairs w_{it}, and demand for food $x^*_{\theta,1}(\cdot)$ is weakly increasing in a univariate parameter α characterizing unobserved heterogeneity, then, loosely speaking, every conditional quantile of consumption of food x_{i1t} is a pseudo-demand function (with the correct form), and the τ^{th} conditional quantile of consumption coincides with the pseudo-demand function associated with the τ^{th} conditional quantile of heterogeneity α such that:

$$Q_{x_{i1t}}(\tau|w_{it}) = x^*_{\theta,1}(w_{it}, \alpha_\tau),$$

where $x^*_{\theta,1}(w_{it}, \alpha)$ is the pseudo-demand function associated with the parameter α evaluated at w_{it}, and α_τ is the τ^{th} conditional quantile of heterogeneity α (see Example 1.9 for a concrete example in which α is the preference parameter in a Stone-Geary specification, Matzkin, 2003, and Imbens and Newey, 2009, for variants of this type of result concerning quantile demand, and Blundell, Kristensen, and Matzkin, 2014, and Blundell, Horowitz, and Parey, 2017, for applications of quantile demand). However, there is no reason to believe that average consumption will have this form. Average consumption will rarely adopt this form because the structure of the problem almost always implies that the conditional distribution of the deviations from the conditional mean is *asymmetric* at some admissible pair $w \in \mathbb{R}^2_{++}$. This asymmetry follows from the fact that, when preferences are stochastic, the locations of the kinks of pseudo-demand $x^*_{\theta,1}(\cdot)$ can vary significantly across households. In such an environment, if we were to average consumption, we could obtain a function with many kinks, or even a function that is strictly increasing, without any kinks (because it is the conditional average of many piecewise linear functions).

Example 1.9. Suppose that the household has a Stone-Geary utility function, defined by: $u(x) = x_1^\alpha x_2^{1-\alpha}$, for every $x \in \bar{R}$. Under this specification, standard demand for food has the form: $x_{u,1}(z) =$

[9] Formally, I define $Q_{x_{i1t}}(\tau|w_{it})$ to be the largest value of q such that $\Pr(x_{i1t} \leq q|w_{it}) = \tau$.

$\alpha y/p$, for each $z \in \mathbb{R}^2_{++}$. Further suppose that the policy $b(\cdot)$ has the form with fixed prices in (1.2.6) subject to $\gamma_2 < 1$. These assumptions imply that the pseudo-demand function $x^*_{\theta,1}(\cdot, \alpha)$ has the form:

$$x^*_{\theta,1}(w,\alpha) = \begin{cases} \frac{\alpha e}{p}, & \text{if } e > \frac{\gamma_1}{\alpha+\gamma_2(1-\alpha)}, \\ \frac{\gamma_1-\gamma_2 e}{p(1-\gamma_2)}, & \text{if } \frac{\gamma_1(\alpha+\pi(1-\alpha))}{\alpha+\pi\gamma_2(1-\alpha)} \leq e \leq \frac{\gamma_1}{\alpha+\gamma_2(1-\alpha)}, \\ \frac{\alpha e}{p(\alpha+\pi(1-\alpha))}, & \text{if } \gamma_1(\alpha+\pi(1-\alpha)) < e < \frac{\gamma_1(\alpha+\pi(1-\alpha))}{\alpha+\pi\gamma_2(1-\alpha)}, \end{cases} \tag{1.4.14}$$

for every admissible $w \in \mathbb{R}^2_{++}$. The form of this function can be verified by following the usual steps: constructing demand $x_\theta(\cdot, \alpha)$ and expenditure $e_\theta(\cdot, \alpha)$, inverting expenditure $e_\theta(\cdot, \alpha)$ to get the pseudo-income function $y^*_\theta(\cdot, \alpha)$, then plugging this function into demand $x_\theta(\cdot, \alpha)$. These steps are omitted for brevity. It is also easy to verify that this function coincides with the first component of the pseudo-demand function $x^*_\theta(\cdot)$ in (1.2.40) when $\alpha = 1/2$, as expected. Now, let us suppose that the household's relative preference for food α is independently drawn from a uniform distribution with support $[1/4, 3/4]$. We can immediately see that the lowest admissible value of expenditure e, given p, varies with α, and that this lower bound ranges from $\frac{\gamma_1(1+3\pi)}{4}$ to $\frac{\gamma_1(3+\pi)}{4}$. These bounds do not depend on p because, under this specification, the locations of the boundaries of the regimes of pseudo-demand do not depend on p (see the form in (1.4.14), and Figure 1.9 for these boundaries given $\alpha = 1/2$). Since $x^*_{\theta,1}(\cdot, \alpha)$ is weakly increasing in α, the τ^{th} conditional quantile for consumption of food x_{i1t} coincides with the pseudo-demand function associated with the τ^{th} quantile of the distribution of α (although not uniquely when it is in the second regime), at every $e \geq \frac{\gamma_1(3+\pi)}{4}$. In particular, it coincides with $\alpha = \frac{1+2\tau}{4}$. We also see that, for any admissible pair w, if the pseudo-demand function associated with α is in the third regime, then so is the pseudo-demand function associated with any $\alpha' \leq \alpha$, verfiying the idea that fraud is "more likely" to occur at lower quantiles. I illustrate the conditional mean, and two conditional quantiles in Figure 1.16. In this figure, average consumption does not have the form of a pseudo-demand function (as it appears continuously-differentiable), but the quantiles do whenever e is larger than $\frac{\gamma_1(3+\pi)}{4}$. If we were to introduce a positive probability that the household does not receive benefits, even though it is eligible, and a positive probability that the household does not commit fraud, even though it has an incentive to do so, we would lose the equivalence between the conditional quantiles and pseudo-demand, but, if these "errors" are independent of the distribution of heterogeneity α in the population, then there would still exist a value $\tau \in (0, 1)$ such that $Q_{x_{i1t}}(\tau|w_{it})$ coincides with the pseudo-demand function $x^*_{\theta,1}(\cdot, \alpha_\tau)$. △

1.4.3 Non-Parametric Estimation

The assumptions on the observations and latent stochastic model in Section 1.4.2 provide the structure that we need to discuss the non-parametric estimation of pseudo-demand $x^*_\theta(\cdot)$. The primary difficulty follows from the fact that this function is non-differentiable along two curves on the interior of the set of admissible pairs $w \in \mathbb{R}^2_{++}$, and that these curves have unknown locations. These curves are the *ridge*

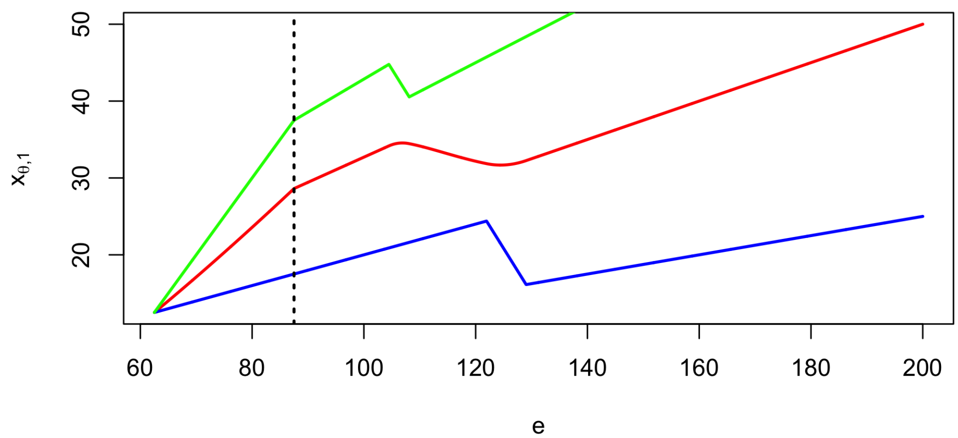

Figure 1.16. **Average and Quantile Demand in Example 1.9.** The red line denotes the conditional average of consumption for food. The blue line denotes the quantile associated with $\tau = 0$. The green line denotes the quantile associated with $\tau = 1$. The dotted line denotes $\frac{\gamma_1(3+\pi)}{4}$. This figure was generated using $\gamma_1 = 100$, $\gamma_2 = 0.7$, and $\pi = 0.5$.

and *valley* curves, respectively characterizing the "lower" and "upper" boundaries of the second regime, as defined in Section 1.2.12. In this section, I develop a non-parametric estimator for pseudo-demand $x^*_\theta(\cdot)$ and illustrate its performance using Monte Carlo simulations.

I consider two steps: First, I approximate the pseudo-demand function $x^*_{\theta,1}(\cdot)$ using linear splines in a quantile regression with a least absolute shrinkage and selection operator (LASSO). Since pseudo-demand $x^*_{\theta,1}(\cdot)$ is continuously-differentiable everywhere except the curves of interest, the partial derivative of a good approximation of pseudo-demand with respect to expenditure e will experience a large "jump" at these curves. Since pseudo-demand $x^*_{\theta,1}(\cdot)$ is strictly increasing in the first and third regimes, and decreasing in the second regime, this "jump" will always be negative on the ridge curve, and will always be positive on the valley curve. Second, I estimate the ridge and valley curves by solving for the function of price p on which the approximation of pseudo-demand $x^*_{\theta,1}(\cdot)$ has (i) the largest *negative* jump in its partial derivative with respect to e, and (ii) the function of price p on which the approximation of pseudo-demand $x^*_{\theta,1}(\cdot)$ has the largest *positive* jump in its partial derivative with respect to e. Since the form of our spline approximation is known, these functions have closed-form solutions. After estimating the ridge and valley curves, the econometrician could go back to improve the estimate of pseudo-demand $x^*_{\theta,1}(\cdot)$ by introducing a constrained non-parametric estimator. I do not complete this additional step because the spline estimator performs well.

Spline Approximation of Pseudo-Demand

Let us start by describing how to estimate pseudo-demand $x^*_{\theta,1}(\cdot)$ without knowing the location of the non-differentiable curves. Formally, I use the following linear spline approximation:

$$\begin{aligned} s(w;\mu,\kappa,\xi) &\equiv \beta_0 + \beta_1 e + \beta_2 p + \beta_3 ep \\ &+ \sum_j \delta_j (e-\kappa_j)^+ + \sum_k \nu_k (p-\xi_k)^+ + \sum_{j,k} \eta_{j,k}(e-\kappa_j)^+(p-\xi_k)^+, \end{aligned} \tag{1.4.15}$$

where $\kappa_1 < \cdots < \kappa_J$ denotes a collection of "knots" for expenditure, $\xi_1 < \cdots < \xi_K$ denotes a collection of "knots" for price, and $\mu = (\beta,\delta,\nu,\eta) \in \mathbb{R}^{4+J+K+JK}$ denotes a vector of real-valued coefficients (see Chen, 1993, Stone, 1994, and Chen, 1997, for the use of products in multivariate splines, and Section 2.4 in Harrell, 2001, for an overview of spline approximations). The spline function $s(\cdot)$ approximates pseudo-demand $x^*_{\theta,1}(\cdot)$ by gluing together small planes. It is continuous because it is the sum of continuous functions. Clearly, the set of all linear splines with the form in (1.4.15) is dense in the set of all continuous real-valued functions on compact sets with respect to the topology of uniform convergence. Therefore, for any pseudo-demand function $x^*_{\theta,1}(\cdot)$, there exists a collection of knots and parameters for which $s(w;\mu,\kappa,\xi)$ is arbitrarily close to $x^*_{\theta,1}(\cdot)$ with respect to the uniform norm over the set $[\kappa_1,\kappa_J]\times[\xi_1,\xi_K]$.

Because the linear spline approximation in (1.4.15) has $4+J+K+JK$ parameters, and we want J and K to be large (to ensure a good approximation), if we were to estimate this approximation by minimizing the sum of least absolute deviations, we would risk overfitting our data (see Koenker and Bassett, 1978, for a seminal paper describing quantile regressions, Koenker et al., 1994, for the use of splines in a quantile regression, and Koenker, 2005, for a broad presentation). It is, therefore, common to introduce a penalization (e.g., LASSO). Note, it is best to penalize δ, ν, and η with different weights because the degree of overfitting can vary across these parameters (see Ruppert and Carroll, 1997, for more on penalization).

A lasso-penalized quantile spline estimator minimizes:

$$\sum_{i,t} \rho_\tau\left[x_{i1t} - s(w_{it};\mu,\kappa,\xi)\right] + \lambda_0 \sum_j \delta_j + \lambda_1 \sum_k \nu_k + \lambda_2 \sum_{j,k} \eta_{j,k}, \tag{1.4.16}$$

with respect to the vector of parameters μ given λ, κ, and ξ, in which $\rho_\tau(\cdot)$ is the usual *check function*, defined by $\rho_\tau(q) = q(\tau - \mathbb{1}\{q<0\})$, for every $q \in \mathbb{R}$, and $\lambda \in \mathbb{R}^3_{++}$ is a vector of tuning parameters characterizing the degrees of penalization for δ, ν, and η (see Tibshirani, 1996, for more on LASSO). Let $\hat{\mu}_\tau(\kappa,\xi)$ denote the argument that minimizes (1.4.16) given the parameter λ and the knots κ and ξ.

The estimator $\hat{\mu}_\tau(\kappa,\xi)$ is *parametric* because it considers the number and location of knots to be fixed. We can construct a *non-parametric* estimator by allowing the number of knots to increase with the number of observations nT. Loosely speaking, if we increase the number of knots, it is important to ensure that these knots become sufficiently "dense" in the domain of interest. In what follows, I fix a closed rectangle of admissible pairs $w \in \mathbb{R}^2_{++}$ by fixing the end points of the domain of interest—that is, κ_1, κ_J, ξ_1, and ξ_K. I, then, take a grid of evenly spaced points over this rectangle, and estimate pseudo-demand $x^*_{\theta,1}(\cdot)$ over this interval by allowing the number of evenly spaced points to tend to infinity. Because smoothing is moderated with penalization, the number of knots is not too important, as long as K and J are above minimum thresholds, and increasing in nT (see, for instance, the introduction in Ruppert, 2002, for an overview of the literature on knot selection, and Section 3 in the same paper for a discussion). One simple option is to choose the number of knots such that $K = J$ using a cross-validation procedure by applying Akaike's Information Criterion (AIC).

Remark 1.11. In general, there is no closed-form solution to the minimization problem associated with (1.4.16). Furthermore, when J and K are large, it is computationally infeasible to optimize (1.4.16). Fortunately, we can solve this problem by reformulating the problem as a linear program. I make use of the modified version of the Barrodale and Roberts (1974) algorithm, defined in Koenker and D'Orey (1987, 1994) and implemented in the *quantreg* package in R. To illustrate, let us consider the median. In this case, the Barrodale and Roberts (1974) algorithm solves the problem:

$$\begin{gathered}\min \sum_{i,t} u_{it} + v_{it} \;\text{ s.t. }\; s(w_{it};\mu_0,\kappa,\xi) - s(w_{it};\mu_1,\kappa,\xi) + u_{it} - v_{it} = y_{it}, \\ u_{it} \geq 0,\;\; v_{it} \geq 0,\;\; \mu_0 \geq 0,\;\; \text{and}\;\; \mu_1 \geq 0,\;\; \forall i,t,\end{gathered} \tag{1.4.17}$$

with respect to u_{it}, v_{it}, μ_0, and μ_1. This reformulation separates the absolute value in the objective function in (1.4.16) into its positive and negative components, removing the non-linearity in this objective function (a procedure sometimes referred to as variable splitting), and separates the parameter μ into its positive and negative components, so that we can apply standard efficient linear programming methods.

Ridge and Valley Estimation

We can now use the spline approximation of pseudo-demand $x^*_{\theta,1}(\cdot)$ to construct a non-parametric estimator for the ridge and valley of pseudo-demand $x^*_{\theta,1}(\cdot)$. As mentioned, we are looking for the locations of negative and positive jumps in the partial derivative of the approximation of the pseudo-demand function $x^*_{\theta,1}(\cdot)$ with respect to expenditure e. Indeed, pseudo-demand $x^*_{\theta,1}(\cdot)$ is continuously-differentiable everywhere except the curves that we want to estimate, implying that its left and right partial derivatives with respect to expenditure should align everywhere except on these curves. The procedure in this section is related to those in Mueller (1992), Mueller and Song (1997), and Huh and Carriere (2002), and the "diagnostic step" in Gijbels and Goderniaux (2004b), the last of which makes use of the second derivative (see Klotsche and Gloster, 2012, for a review of some techniques used to non-parametrically estimate the location of a "kink" in a univariate setting). There are four primary differences between the estimator in this section and most of this body of literature: (i) the estimator in this section is constructed with the specific intention of detecting *more than one kink*, (ii) the estimator in this section looks for one kink that is known to have an *increase* in the partial derivative of pseudo-demand $x^*_{\theta,1}(\cdot)$ with respect to expenditure, and one kink that is known to have a *decrease* in the partial derivative of pseudo-demand $x^*_{\theta,1}(\cdot)$ with respect to expenditure (rather than, say, two kinks with unknown directions for the changes in this partial derivative), (iii) the estimator in this section is applied to the *conditional quantile* (rather than the mean), and, most importantly, (iv) the estimator in this section is for a bivariate environment (instead of a univariate environment), so that we are estimating non-parametric *functions*, not points.

The partial derivative of the spline approximation $s(w;\mu,\kappa,\xi)$ is equal to:

$$\frac{\partial s(w;\mu,\kappa,\xi)}{\partial e} = \beta_1 + \beta_3 p + \sum_j \delta_j \mathbb{1}\{e > \kappa_j\} + \sum_{j,k} \eta_{j,k} \mathbb{1}\{e > \kappa_j\}(p - \xi_k)^+, \tag{1.4.18}$$

at every admissible $w \in \mathbb{R}^2_{++}$ such that $e \neq \kappa_j$, for any $j = 1, \ldots, J$, where $\mathbb{1}\{\cdot\}$ is the indicator function that is equal to 1 if the logical argument inside the brackets is true, and equal to 0, otherwise. Therefore, the left and right partial derivatives of this approximation are equal whenever expenditure e is not at a knot κ_j. Conversely, this partial derivative jumps at every knot κ_j, and if there exists a point with a large jump in the partial derivative of this approximation, it necessarily happens at one of the knots. The difference in the left and right partial derivatives at knot κ_j is:

$$\Delta_j(p;\mu,\kappa,\xi) = \delta_j + \sum_k \eta_{j,k}(p - \xi_k)^+. \tag{1.4.19}$$

If the number and location of knots were fixed, then a natural estimator for the ridge curve $r(p)$ would be the knot κ_j with the lowest $\Delta_j(p;\hat{\mu}_\tau(\kappa,\xi),\kappa,\xi)$ at each p, and a natural estimator for the valley curve $v(p)$ would be the knot κ_j with the highest $\Delta_j(p;\hat{\mu}_\tau(\kappa,\xi),\kappa,\xi)$ at each p. The resulting estimators are necessarily step functions with codomain $\{\kappa_1,\ldots,\kappa_J\}$. These estimators can be extended to the setting in which knots are not fixed (allowing the number of knots to grow with nT). Let $\hat{r}(p)$ and $\hat{v}(p)$ denote these estimators, respectively. While omitted, these estimators can be smoothed after the fact, if desired.

Discussion

Let us now discuss some aspects of the estimators described above. First, it is worth mentioning that there is no particular reason why I use a *linear* spline approximation instead of, say, a cubic spline approximation. One advantage of the linear spline is the resulting simplification of the optimization problem used to find the location of the smallest or largest jump in the partial derivative of the approximation with respect to expenditure. Second, notice that, instead of using splines, I could have used another non-parametric estimator in the first step. For example, I could have used a local linear (or quadratic) fit to estimate a smooth approximation of pseudo-demand $x^*_{\theta,1}(\cdot)$ (see Nadaraya, 1964, and Watson, 1964, for seminal papers on local regressions, Lu, 1996, for local linear and quadratic fits, as well as Yu and Jones, 1998, and Hallin et al., 2009, for the application of such estimators to quantile regressions). There are two issues with this alternative: (i) It would make it computationally challenging to find the ridge and valley curves, and (ii) it is *local* (unlike the spline approximation), implying that the asymptotics for the resulting estimators for the ridge and valley curves would have to be evaluated pointwise, which is not desirable. Another example would be a polynomial sieve estimator. This type of estimator would fix the problems associated with the local estimator described above, but it is often unstable at the boundaries, making it more likely to "miss" the ridge and valley curves (see Section 2.4.8 in Harrell, 2001, for a broad discussion of the advantages of splines). Third, there is a body of statistical literature addressing the optimal estimation of change-points in the derivatives of univariate functions observed with error (see Cheng and Raimondo, 2008, Wishart, 2010, 2011a, 2011b, Han et al., 2014, Bengs and Holzmann, 2019, and Tuvaandorj, 2020). This type of analysis is, however, challenging to extend to a bivariate environment—the estimator from the previous section is likely the best place to start. Fourth, it is worth mentioning that, for the two-step estimator described in this section to work, it is vital to have a grid with enough knots. If we have only a few knots and the ridge or valley curve lies between two knots, then the jump in the derivative that we are looking for can be averaged between the knots above and below this curve, making it unlikely that we will have the ability to detect this curve.[10] Fifth, it is always possible to find a ridge and valley using the estimator described in this section (although they might be small or have the wrong order). Tests for kinks exist in similar environments (see Gijbels and Goderniaux, 2004a), but a test for the current setting is left for future research.

Monte Carlo Simulations

I now provide Monte Carlo simulations to illustrate the small sample performance of the two-step estimator for the ridge and valley curves described above. To simulate, I assume that preferences have the Stone-Geary specification in Example 1.9, and provide results for two distinct data generating processes:

(i) **First data generating process:** I draw income y from a log-normal distribution with mean 4 and standard deviation 0.5 (on the log-scale), and draw the price p from a log-normal distribution with mean 0 and standard deviation 0.5 (on the log-scale). I draw a preference parameter α from a uniform distribution with support $[1/4, 3/4]$. I construct expenditure and pseudo-demand given α, γ, and π, as defined in (1.4.14). Consumption x_{i1t} equals pseudo-demand for food.

(ii) **Second data generating process:** I draw income y from a log-normal distribution with mean 4 and standard deviation 0.5 (on the log-scale), and draw the price p from a log-normal distribution with mean 0 and standard deviation 0.5 (on the log-scale). I construct expenditure and pseudo-demand given $\alpha = 0.5$, γ, and π, as defined in (1.4.14). I construct consumption x_{i1t} for food by adding an error to pseudo-demand for food. This error is drawn from a normal distribution with

[10] I also recommend using evenly-spaced knots because our main concern is having enough knots around the locations of the ridge and valley curves (which are not known ex ante). This practice differs from the standard practice of placing the knots at the quantiles of the design variables—that is, e and p.

mean 0 and standard deviation 3.

In the first data generating process, preferences are stochastic; in the second data generating process, preferences are fixed, but consumption is observed with error. For these simulations, I use $\gamma_1 = 100$, $\gamma_2 = 0.7$, and $\pi = 0.5$, and I focus on the conditional median, for simplicity. In the first data generating process, the conditional mean of consumption does not coincide with a pseudo-demand function, as described in Section 1.4.2 and illustrated in Example 1.9; in the second data generating process, the conditional mean and median coincide. In each case, the conditional median of consumption coincides with the pseudo-demand function in (1.4.14) given $\alpha = 0.5$, at every expenditure e larger than $\frac{\gamma_1(3+\pi)}{4} = 87.5$. This pseudo-demand function is illustrated in Figure 1.17. In each case, the ridge curve $r(p) = \frac{(1+\pi)\gamma_1}{1+\pi\gamma_2}$ is constant at approximately 111.11, and the valley curve $v(p) = \frac{2\gamma_1}{1+\gamma_2}$ is constant at approximately 117.64. A simulated sample distribution of expenditures e_{it} and prices p_{it} is illustrated in Figure 1.18. The colour of each observation in this figure denotes the value of consumption x_{i1t}. Notice that, it is difficult to visually discern the non-linearity in consumption using the naked eye. This figure is intended to illustrate that fraud can be "hidden," and that we require the proper economic tools to be able to determine whether or not there exists fraud in a dataset. For each data generating process, I approximate the pseudo-demand function using the spline estimator described in Section 1.4.3, then estimate the ridge and valley curves using the procedure in Section 1.4.3. In Tables 1.1 and 1.2, I provide the results for $nT = 500$ and $nT = 1000$.[11] Each table provides the mean and variance of several measures of distance between the true ridge and valley curves, and their estimated counterparts. Knots are evenly spaced on $[90, 135] \times [0.5, 1.5]$. This domain is chosen to ensure that observations are well-distributed (see the samples in Figure 1.18). When $nT = 500$, I use $J = K = 50$, $\lambda_0 = 50$, $\lambda_1 = 110$, and $\lambda_2 = 35$, and when $nT = 1000$, I use $J = K = 70$, $\lambda_0 = 40$, $\lambda_1 = 100$, and $\lambda_2 = 25$. Results are calculated from $S = 100$ draws. In each table, we see that $\hat{r}(p)$ underestimates $r(p)$, and that $\hat{v}(p)$ overestimates $v(p)$. This bias follows from the fact that we are smoothing the pseudo-demand function in our approximation. Fortunately, this bias reduces quickly with the sample size nT. The variance of each measure of distance also decreases quickly in the sample size nT. In addition to the numbers reported in these tables, it should be noted that, in each simulation (consisting of 100 draws and 50 points of evaluation for both $\hat{r}(p)$ and $\hat{v}(p)$ in each draw), there were less than 6 instances (or $\frac{6\times 100}{2\times 50\times 100} = 0.06\%$ of evaluations) in which an estimate of a ridge or valley touched the boundary of the grid. Since this percent is small, we can conclude that the size of the chosen rectangle $[90, 135] \times [0.5, 1.5]$ had a negligible impact on these results. On the left of Figure 1.19, I illustrate the spline approximation of pseudo-demand $x^*_{\theta,1}(\cdot)$ for one simulated sample given the first data generating process. As expected, this spline approximation looks similar to the pseudo-demand function in Figure 1.17. On the right of Figure 1.19, I illustrate an estimate of the ridge and valley curves. In this figure, we see three important features: (i) the estimated curves are close to the true curves, (ii) the estimated curves are step functions, and (iii) $\hat{r}(p)$ is below

[11] Note, under Assumption 1.9(i), estimation depends on the sample size nT, but not separately on the value of n or T.

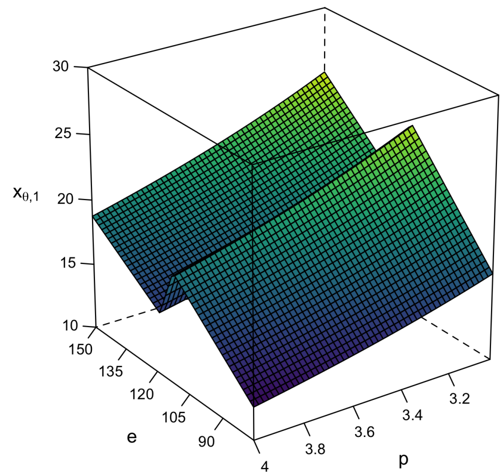

Figure 1.17. Pseudo-demand $x^*_{\theta,1}$ in (1.2.40) given $\gamma_1 = 100$, $\gamma_2 = 0.7$, and $\pi = 0.5$.

$r(p)$ and $\hat{v}(p)$ is above $v(p)$. The third feature is expected given the bias observed in Tables 1.1 and 1.2.

1.5 Application

In this section, I apply the economic tools from Sections 1.2 to 1.4. First, I analyze the Panel Survey of Income Dynamics (PSID). Second, I analyze the Nielsen Homescan Consumer Panel (NHCP). The analysis of the PSID aids in the analysis of the NHCP.

1.5.1 Panel Survey of Income Dynamics

The Panel Survey of Income Dynamics (PSID) is a longitudinal survey in the United States. The PSID has collected information on households (and their descendants) since 1968. The original sample consists of approximately 5,000 households. Roughly 60% of this sample is representative of the population in the United States. The remaining 40% of the sample consists of low-income households. The households are surveyed every year from 1968 to 1997, and every second year from 1999 to 2017. The PSID includes questions on household characteristics, income, benefits, and expenditures. In 1999, 2001, and 2003, households were asked whether they have been disqualified from receiving benefits for breaking the rules since the previous survey. In Appendix 1.E, I provide summary statistics and a description of formatting.

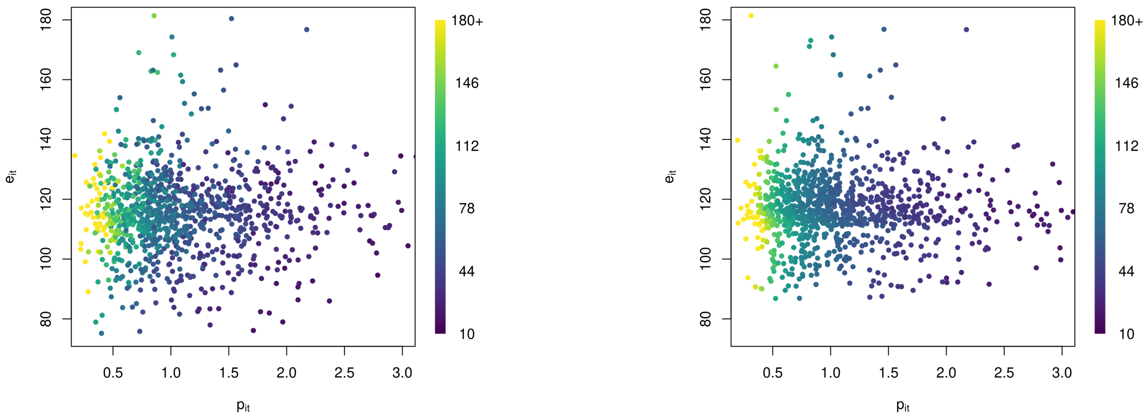

Figure 1.18. **Simulated Samples.** On the left, I illustrate a simulated sample using the first data generating process. On the right, I illustrate a simulated sample using the second data generating process. These figures use the same draws for z. Colour indicates the value of consumption x_{i1t}.

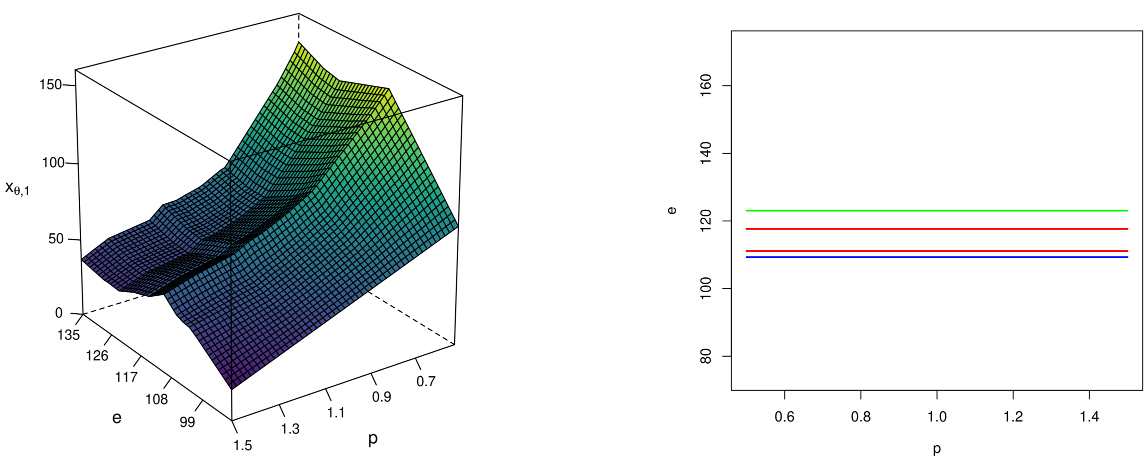

Figure 1.19. **Ridge and Valley Estimates.** On the left, I illustrate an estimated linear spline approximation of the pseudo-demand $x^*_{\theta,1}(\cdot)$ under the first data generating process. On the right, I illustrate the corresponding ridge and valley estimates, $\hat{r}(p)$ and $\hat{v}(p)$. The red lines are $r(p)$ and $v(p)$. The blue line is $\hat{r}(p)$. The green line is $\hat{v}(p)$.

Table 1.1. Monte Carlo results associated with the first data generating process. For these simulations, I use $\gamma_1 = 100$, $\gamma_2 = 0.7$, and $\pi = 0.5$, and estimate the location of the ridge $r(p)$ and valley $v(p)$ for the conditional median at $p = 3, 4$. When $nT = 500$, I use $J = K = 50$, $\lambda_0 = 50$, $\lambda_1 = 110$, and $\lambda_2 = 35$, and when $nT = 1000$, I use $J = K = 70$, $\lambda_0 = 40$, $\lambda_1 = 100$, and $\lambda_2 = 25$. Results are calculated from $S = 100$ draws.

	$nT = 500$		$nT = 1000$	
	Mean	Var.	Mean	Var.
$\hat{r}^s(3) - r(3)$	-1.92	0.74	-1.59	0.36
$\hat{v}^s(3) - v(3)$	7.45	29.52	4.62	26.44
$\hat{r}^s(4) - r(4)$	-1.06	13.54	-1.35	1.89
$\hat{v}^s(4) - v(4)$	6.14	77.79	4.92	29.39
$\frac{1}{K}\sum_k \left(\hat{r}^s(\xi_k) - r(\xi_k)\right)$	-1.41	5.82	-1.44	0.98
$\frac{1}{K}\sum_k \left(\hat{v}^s(\xi_k) - v(\xi_k)\right)$	6.34	66.03	4.76	26.67
$\frac{1}{K}\sum_k \lvert\hat{r}^s(\xi_k) - r(\xi_k)\rvert$	2.31	3.86	1.66	0.55
$\frac{1}{K}\sum_k \lvert\hat{v}^s(\xi_k) - v(\xi_k)\rvert$	8.41	38.11	5.21	22.20
$\left(\int_{0.5}^{1.5} (\hat{r}^s(p) - r(p))^2 dp\right)^{1/2}$	2.42	5.24	1.69	0.63
$\left(\int_{0.5}^{1.5} (\hat{v}^s(p) - v(p))^2 dp\right)^{1/2}$	8.51	39.12	5.31	23.23

Table 1.2. Monte Carlo results associated with the second data generating process. For these simulations, I use $\gamma_1 = 100$, $\gamma_2 = 0.7$, and $\pi = 0.5$, and estimate the location of the ridge $r(p)$ and valley $v(p)$ for the conditional median at $p = 3, 4$. When $nT = 500$, I use $J = K = 50$, $\lambda_0 = 50$, $\lambda_1 = 110$, and $\lambda_2 = 35$, and when $nT = 1000$, I use $J = K = 70$, $\lambda_0 = 40$, $\lambda_1 = 100$, and $\lambda_2 = 25$. Results are calculated from $S = 100$ draws.

	$nT = 500$		$nT = 1000$	
	Mean	Var.	Mean	Var.
$\hat{r}^s(3) - r(3)$	-2.55	23.43	-1.09	2.33
$\hat{v}^s(3) - v(3)$	3.22	12.19	1.27	1.12
$\hat{r}^s(4) - r(4)$	-1.57	54.51	-0.69	18.44
$\hat{v}^s(4) - v(4)$	2.42	15.21	1.23	1.19
$\frac{1}{K}\sum_k \left(\hat{r}^s(\xi_k) - r(\xi_k)\right)$	-2.15	30.90	-0.99	4.20
$\frac{1}{K}\sum_k \left(\hat{v}^s(\xi_k) - v(\xi_k)\right)$	2.87	4.74	1.23	1.00
$\frac{1}{K}\sum_k \lvert\hat{r}^s(\xi_k) - r(\xi_k)\rvert$	3.87	22.52	1.49	2.94
$\frac{1}{K}\sum_k \lvert\hat{v}^s(\xi_k) - v(\xi_k)\rvert$	2.97	4.20	1.31	0.79
$\left(\int_3^4 (\hat{r}^s(p) - r(p))^2 dp\right)^{1/2}$	4.07	26.43	1.60	5.95
$\left(\int_3^4 (\hat{v}^s(p) - v(p))^2 dp\right)^{1/2}$	3.14	6.48	1.34	0.81

Table 1.3. Number and proportion of households that report being disqualified from receiving food stamps for breaking the rules (conditional on receiving a positive amount of benefits or being disqualified) in the PSID by year, in the years in which this question was asked.

Year	Number	Proportion
1999	22	0.0316
2001	7	0.0151
2003	15	0.0171

Food Stamp Fraud

First, let us begin by analyzing the extent of benefit fraud in the PSID. There are two natural ways to approach this analysis: First, we can analyze the proportion of households that report being disqualified from receiving food stamps for breaking the rules (since the previous survey). This preliminary analysis is presented in Table 1.3. In this table, we see that approximately 2 percent of households with benefits report being disqualified in each survey year. This number should be interpreted with some caution: This number likely includes households that have been disqualified for reasons other than the type of fraud described in this chapter. Furthermore, this number only includes households that (i) broke the rules, (ii) got caught, (iii) got disqualified (which requires a sufficient amount of evidence), and (iv) reported it. Second, because we have information on benefits and food expenditure, we can apply Corollary 1.1 to compute an estimate of the amount of fraud in the PSID for households that have not been disqualified. Table 1.4 provides the number and proportion of households that report spending less on food than they receive in benefits. This table also reports the average difference between food expenditure and benefits (conditional on food ependiture being smaller than benefits). In this table, we see that approximately 2 to 10 percent of households that receive benefits have consumption that is consistent with fraud, and that, on average, these households are spending approximately \$325 to \$1,800 less on food than they receive in benefits. In Figure 1.20, we see the distribution of the difference between food expenditure and benefits. These numbers can be used to estimate the expected amount of fraud in the economy. For example, there were approximately 42,123,000 participants in SNAP in 2017 (see Table 1.6 in Appendix 1.A). Therefore, if the probability of fraud is 0.0229 and the expected amount of benefits exchanged conditional on fraud is \$1,108.13, then just over \$1 billion dollars of benefits were illegally exchanged in 2017. This estimate is extremely similar to the estimate of \$1.1 billion in (Willey et al., 2017). Of course, these numbers do not account for misreporting or measurement error. Unfortunately, we cannot learn much more from the PSID because it does not have detailed information about prices or non-food expenditures.

Table 1.4. This table reports the number and proportion of households that report spending less than their allotment of benefits on food (intended for consumption at home), and total food in the PSID by year, and the mean difference between benefits and food expenditure (conditional on being positive). Results are conditional on having positive benefits and expenditure.

	Food at Home			Total Food		
Year	Number	Proportion	Difference	Number	Proportion	Difference
1999	10	0.0184	770.40	31	0.0487	324.29
2001	16	0.0441	1,444.18	25	0.0589	1,294.32
2003	26	0.0388	696.73	46	0.0585	800.78
2005	28	0.0342	782.60	71	0.0703	464.15
2007	40	0.0457	512.92	71	0.0665	434.15
2009	30	0.0350	1,215.46	67	0.0633	891.14
2011	52	0.0460	1,531.53	131	0.0906	926.20
2013	54	0.0387	1,786.07	121	0.0668	1,070.20
2015	35	0.0271	1,601.31	82	0.0537	932.42
2017	30	0.0229	1,108.13	50	0.0328	663.18

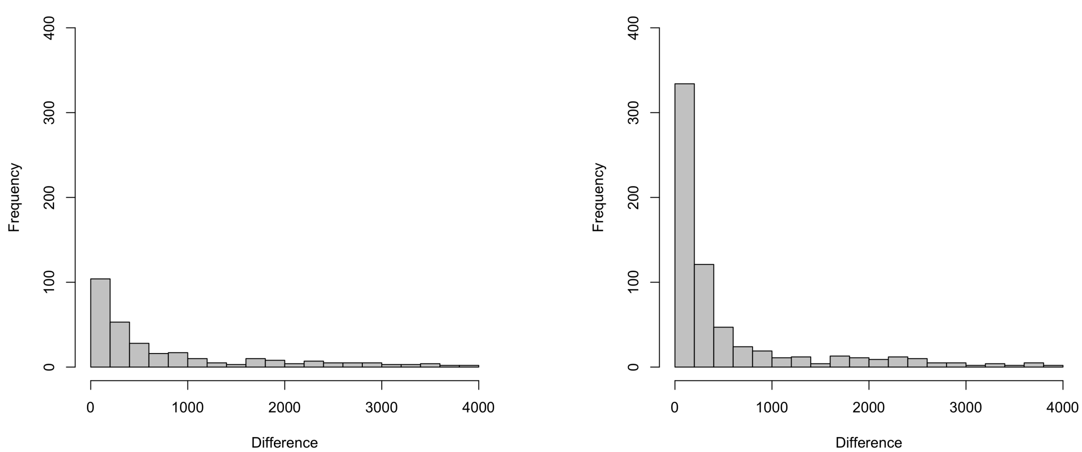

Figure 1.20. **Differences Between Food Expenditure and Benefits.** On the left, I illustrate the distribution of the difference between food expenditure (intended for consumption at home) and benefits in the PSID, conditional on the households spending less on food than it receives in benefits. On the right, I illustrate the distribution of the difference between total food expenditure and benefits in the PSID, conditional on the households spending less on food than it receives in benefits.

Demographics

While the above analysis leaves much to be desired, it is useful because it provides a way to determine the demographics of the households that have been disqualified, and those that exhibit behaviour consistent with benefit fraud. These demographics can, then, be used to construct a subsample in a more sophisticated analysis. In this section, I summarize the important results from a more complete analysis of the demographics of households in Appendix 1.E.1. In particular, I discuss the primary observable differences between (i) all households in the PSID, (ii) households receiving food stamps in the PSID, (iii) disqualified households in the PSID, and (iv) households that report spending less on food (intended for home consumption) than they receive in benefits in the PSID.[12]

This analysis leads to the following findings: On average, households receiving food stamps have (i) more members, (ii) younger and disproportionately female heads, (iii) lower income, and (iv) less education. They are also disproportionately located in the South. Disqualified households, and households that report spending less on food than they receive in benefits, have similar demographics to the entire subsample of households receiving food stamps, with a few exceptions: These households have even younger and more disproportionately female heads, and even lower income. These households are also even more likely to be located in the South. The finding that these households have even lower income agrees with the theoretical finding that only the poorest households have an incentive to commit fraud.[13]

1.5.2 Nielsen Homescan Consumer Panel

Let us now consider a more sophisticated analysis of food stamp fraud. In this section, I introduce the Nielsen Homescan Consumer Panel (NHCP), estimate quantile pseudo-demand functions, and apply the steps in Section 1.3.3 to get bounds for the structural parameters of interest.

The NHCP is a detailed longitudinal dataset that tracks the purchases of households in the United States. Participating households scan all purchased goods with a barcode scanner provided by Nielsen. Prices are entered by the household or linked with retailer data.

I restrict attention to August to October in 2016. These months are consecutive, and avoid holidays (on which consumption might be irregular) such as Independence Day, Christmas Day, and New Year's Eve. The short time frame reduces the possibility of changing tastes or changing product availability. I keep all households with positive expenditure in eight distinct categories of products[14] in all months. After aggregating purchases by month, we are left with 4,807 households and 14,421 observations. In Ap-

[12] I consider food intended for home consumption because households cannot use food stamps to buy "hot foods or hot food products ready for immediate consumption" (see Section 3(k)(1) in the Food and Nutrition Act of 2008 and Appendix 1.A.8 for more).

[13] This analysis was completed independently for disqualified households and households that report spending less on food than they receive in benefits (see Appendix 1.E.1). The results are presented together because each analysis results in a similar conclusion.

[14] These categories are (i) general merchandise (including health and beauty), (ii) dry grocery, (iii) frozen foods, (iv) dairy, (v) deli and packaged meat, (vi) fresh produce and "magnet data" products (consisting of products without Universal Product Codes such as fresh produce), (vii) non-food grocery, and (viii) alcohol.

Table 1.5. Summary of the normalized expenditure e_{it}, normalized price p_{it}, and consumption x_{it} including the mean, standard deviation, ratio of the standard deviation to the mean, and quantiles.

				Quantiles				
Variable	Mean	Std. Dev.	Ratio	0%	25%	50%	75%	100%
e_{it}	2.55	1.40	0.55	0.21	1.58	2.26	3.17	23.79
p_{it}	1.51	0.64	0.42	0.12	1.07	1.42	1.83	6.92
x_{i1t}	1.04	0.53	0.51	0.09	0.67	0.94	1.30	8.77
x_{i2t}	0.99	0.64	0.64	0.07	0.56	0.85	1.25	10.49

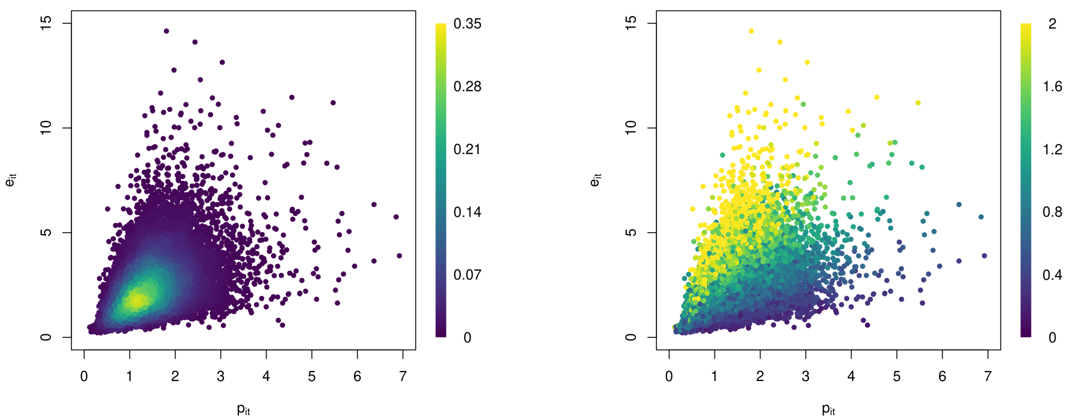

Figure 1.21. Sample distribution of the normalized price p_{it} and normalized expenditure e_{it}. On the left, colour indicates a bivariate kernel density estimate. On the right, colour indicates the quantity of food consumption x_{i1t}.

pendix 1.E.2, I provide a detailed description of variables, formatting, and aggregation (using the method described in Section 1.4.1). I also provide summary statistics and a discussion of the representativeness of the NHCP (after formatting).

In Table 1.5, I provide summary statistics for normalized expenditure e_{it}, the normalized price p_{it}, and consumptions, x_{i1t} and x_{i2t}, pooled across households and months, and in Figure 1.21, I plot the sample distributions of these variables. In particular, on the left of this figure, colour indicates a bivariate kernel density estimate, and on the right of this figure, colour indicates the quantity of food consumption x_{i1t}. In this figure, we see that the joint distribution of (e_{it}, p_{it}) is approximately log-normal, and that there is a lot of heterogeneity in e_{it}, p_{it}, and x_{i1t}.

Estimation of Pseudo-Demand

Under Assumption 1.9, there exists $\tau \in (0, 1)$ for which the conditional quantile $Q_{x_{i1t}}(\tau | w_{it})$ is a pseudo-demand function $x^*_{\theta,1}(\cdot)$. In this section, I investigate (i) whether this assumption holds in the NHCP, and (ii) the quantile(s) at which it holds (when it holds). I look for ridges and valleys in various conditional

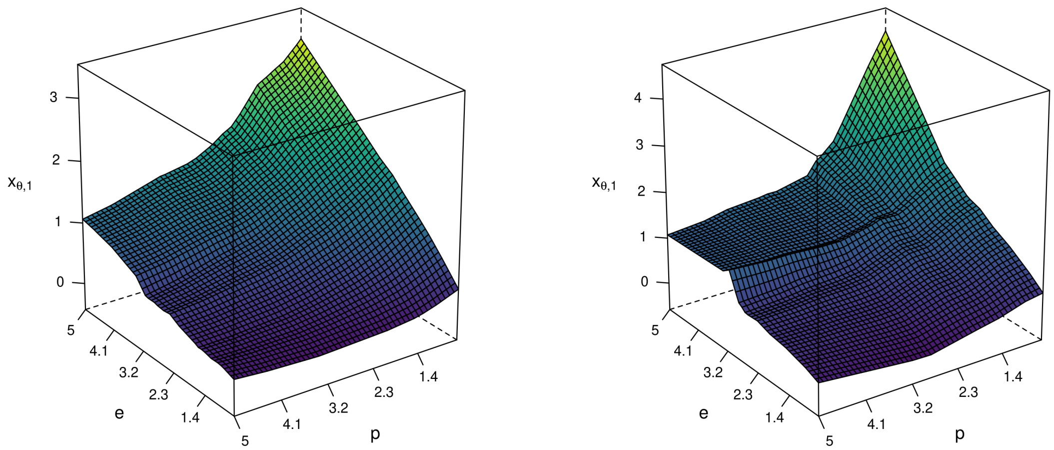

Figure 1.22. Estimates of conditional quantiles of food consumption given $\tau = 0.25$. On the left, I plot the estimate for the full NHCP sample. On the right, I plot the estimate for the subset of the NHCP containing poorer households with more members and younger female heads.

quantiles using the spline approach, and perform this analysis on (i) the full NHCP sample, and (ii) a subset of the NHCP containing poorer households with more members and younger female heads.

For each sample, I estimate six conditional quantiles of consumption: $\tau = 0.10, 0.25, 0.50, 0.75, 0.90$. For brevity, I will only plot the pseudo-demand functions for $\tau = 0.25$ and $\tau = 0.50$ in this section. The remaining plots are placed in Appendix 1.F. This appendix also contains the details of the subset of the NHCP described above, and the tuning parameters used for estimation.

Figures 1.22 and 1.23 display the estimated conditional quantiles of consumption for $\tau = 0.25$ (first quartile) and $\tau = 0.50$ (median), respectively. In each figure, the estimate on the left uses the full NHCP sample, and the estimate on the right uses the subset of the NHCP containing poorer households with more members and younger female heads. In Figure 1.22, the estimate that uses the full sample has a small "wrinkle" when expenditure e is between 3 and 4 in normalized dollars, but this "wrinkle" does not seem to define a very recognizable ridge or valley. However, the estimate that uses the subset of the NHCP has an extremely pronounced ridge and valley in this location. We also see a similar ridge and valley in this subset of the NHCP in Figure 1.23 when expenditure e is between 3.5 and 4.5 in normalized dollars. These results suggest the existence of fraud in this subset of the NHCP. None of the other conditional quantiles in either sample have a very pronounced ridge or valley (see Appendix 1.F).

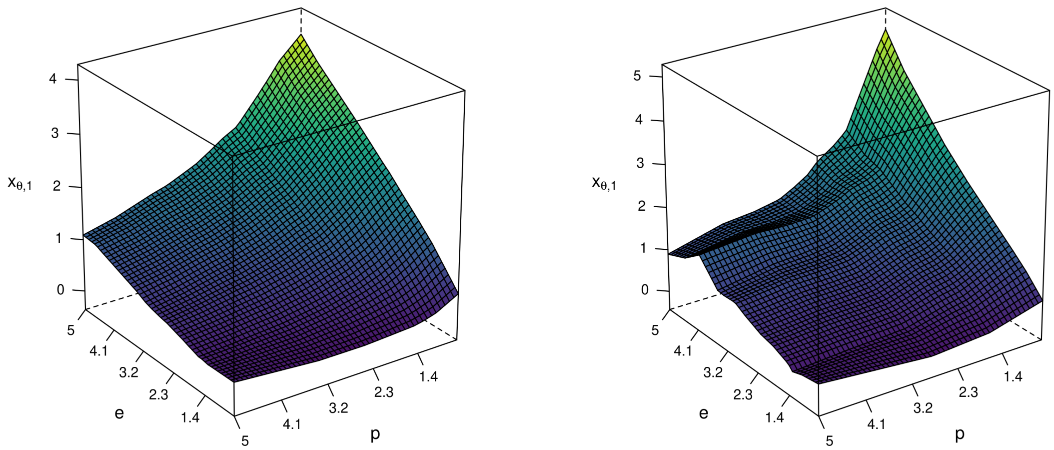

Figure 1.23. Estimates of conditional quantiles of food consumption given $\tau = 0.50$. On the left, I plot the estimate for the full NHCP sample. On the right, I plot the estimate for the subset of the NHCP containing poorer households with more members and younger female heads.

Structural Estimates

The estimates of pseudo-demand on the right of Figures 1.22 and 1.23 can be used to bound the policy $b(\cdot)$, discount π, and demand for fraud $f_\theta(\cdot)$. In this section, I will apply the steps described in Section 1.3.3 to the estimate associated with the subset of the NHCP consisting of poorer households with more members and younger female heads given $\tau = 0.25$.

To identify the regimes, I apply the ridge and valley estimation procedure described in Section 1.4.3. The result of this procedure is displayed in Figure 1.24. In this figure, we see that the order of the ridge and valley is correct at prices that are smaller than 3.25. This result follows from the fact that the valley is not as recognizable at large prices. For simplicity, I use the following estimates for the sets associated with the regimes:

$$\hat{W}^*_{\theta,1} = [4.47, 7] \times [1.05, 3.25], \quad \hat{W}^*_{\theta,2} = [4.21, 4.47] \times [1.05, 3.25], \quad \hat{W}^*_{\theta,3} = [0, 4.21] \times [1.05, 3.25], \quad (1.5.1)$$

avoiding extremal values of the price p. Since these estimated sets are rectangles, we immediately obtain:

$$\hat{e}_\ell(p) = \inf\{e > 0 : (e,p) \in \hat{W}^*_{\theta,2}\} = 4.21 \text{ and } \hat{e}_h(p) = \sup\{e > 0 : (e,p) \in \hat{W}^*_{\theta,2}\} = 4.47, \quad (1.5.2)$$

for all $1.05 \leq p \leq 3.25$.

The next step involves estimating $b^*_\theta(e_\ell(p), p)$ and $b^*_\theta(e_h(p), p)$, and then using the result to estimate

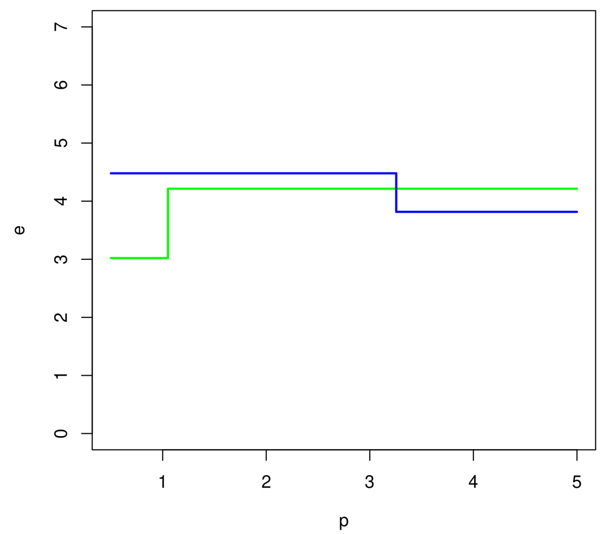

Figure 1.24. Estimates of the ridge and valley for the conditional quantile of food consumption given $\tau = 0.25$. The green curve is the estimated ridge and the blue curve is the estimated valley.

the bounds for $b(\cdot)$ and the upper bound π_h^* for the discount π. We have the following natural estimators:

$$\hat{b}_\theta^*(\hat{e}_\ell(p), p) = p\hat{x}_{\theta,1}^*(\hat{e}_\ell(p), p) \text{ and } \hat{b}_\theta^*(\hat{e}_h(p), p) = p\hat{x}_{\theta,1}^*(\hat{e}_h(p), p), \tag{1.5.3}$$

for all $1.05 \leq p \leq 3.25$. The resulting estimates are illustrated on the left of Figure 1.25. In this figure, we see that these estimates are increasing in the price, and that the estimate associated with the smaller level of expenditure is mostly above the estimate associated with the higher level of expenditure, as expected. The bounds for $b(\cdot)$ implied by these estimates are illustrated on the right of Figure 1.25. These bounds have the expected form. Moreover, we obtain:

$$\hat{\pi}_h^* = \inf_{w \in \hat{W}_{\theta,3}^*} \frac{\hat{x}_{\theta,2}^*(w)}{\hat{b}_\theta^*(\hat{e}_\ell(p), p) - p\hat{x}_{\theta,1}^*(w)} = \inf_{w \in \hat{W}_{\theta,3}^*} \frac{e - p\hat{x}_{\theta,1}^*(w)}{p\hat{x}_{\theta,1}^*(\hat{e}_\ell(p), p) - p\hat{x}_{\theta,1}^*(w)} \simeq 0.1145. \tag{1.5.4}$$

This upper bound is less than one, implying that we can also bound demand for fraud. This upper bound is much lower than the informal claim that the discount π is often around 0.50 (see Government Accountability Office, 2006).

It is left to bound demand for fraud. First, I consider pseudo-demand for fraud $f_\theta^*(\cdot)$, then I consider demand for fraud $f_\theta(\cdot)$ as a function of z. The estimated bounds for pseudo-demand for fraud have the form in Section 1.3.3:

$$\hat{b}_\theta^*(\hat{e}_\ell(p), p) - p\hat{x}_{\theta,1}^*(w) \leq f_\theta^*(w) \leq \hat{b}_\theta^*(\hat{e}_\ell(p), p) + \frac{|\hat{e}_\ell(p) - e|}{1 - \hat{\pi}_h^*} - p\hat{x}_{\theta,1}^*(w), \tag{1.5.5}$$

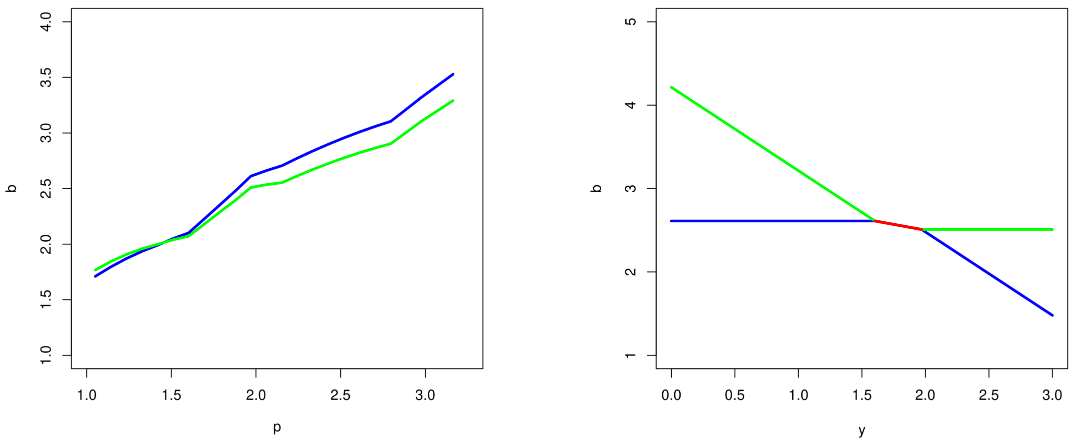

Figure 1.25. On the left, the blue line is the estimate of $b^*_\theta(e_\ell(p), p)$, and the green line is the estimate of $b^*_\theta(e_\ell(p), p)$. On the right, the blue line is the estimate of the lower bound for $b(\cdot)$ given $p = 2$, the green line is the estimate of the upper bound for $b(\cdot)$ given $p = 2$, and the red line is the estimate of $b(\cdot)$ in the second regime given $p = 2$.

for all $w \in \hat{W}^*_{\theta,3}$. These bounds are illustrated in Figure 1.26. In this figure, we see that both bounds are informative, and that they converge to zero at the boundary of the third regime. Now, to estimate the bounds for $f_\theta(\cdot)$, we first need to estimate $\lambda_\ell(z)$ and $\lambda_h(z)$, as defined in Section 1.3.3. I use the following plug-in estimators:

$$\hat{\lambda}_\ell(z) = \{e : \hat{x}_{\theta,1}(w) = y\} \;\text{ and }\; \hat{\lambda}_h(z) = (1 - \hat{\pi}^*_h)\big[y - \hat{x}^*_{\theta,2}(\hat{e}_\ell(p), p)\big] + \hat{e}_\ell(p), \tag{1.5.6}$$

in the third regime. These functions bound the expenditure function $e_\theta(\cdot)$ in the third regime. I illustrate their estimates on the left of Figure 1.27. These estimates are strictly increasing, meet at the boundary of the third regime, and cross at the boundary of the first regime, as expected. Now, the bounds for $f_\theta(\cdot)$ can be estimated in the following way:

$$\hat{b}^*_\theta(\hat{e}_\ell(p), p) - p\hat{x}^*_{\theta,1}(\hat{\lambda}_h(z), p) \leq f_\theta(z) \leq \hat{b}^*_\theta(\hat{e}_\ell(p), p) + \frac{\hat{e}_\ell(p) - \hat{\lambda}_\ell(z)}{1 - \hat{\pi}^*_h} - p\hat{x}^*_{\theta,1}(\hat{\lambda}_\ell(z), p), \tag{1.5.7}$$

in the third regime. The estimated bounds are illustrated on the right of Figure 1.27. Once again, these estimated bounds have the form that we expect: They are strictly decreasing in income and they meet at the boundary of the third regime. These estimators for the pseudo-fraud function $f^*_\theta(\cdot)$ and demand for fraud $f_\theta(\cdot)$ can be used to inform policy: For example, they can be used to (i) determine whether these bounds are changing over time, (ii) rule out fraud for households with sufficiently high levels of income or expenditure, or (iii) bound the effect of a change in policy on fraud.

Remark 1.12. If we have the interpretation that this first quartile of food consumption $Q_{x_{i1t}}(0.25|w_{it})$

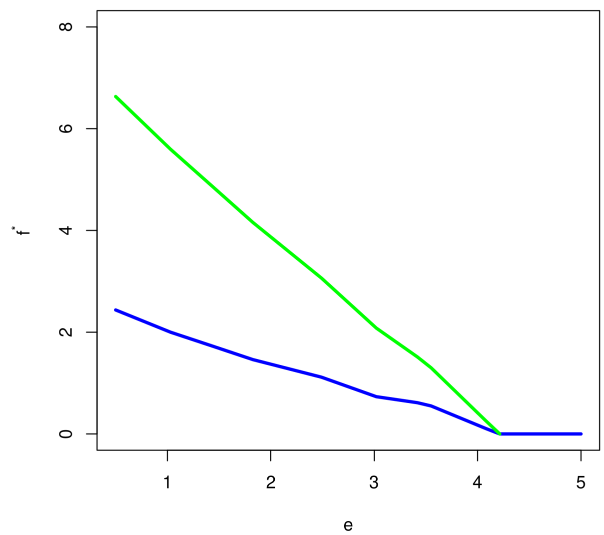

Figure 1.26. The estimated bounds for the pseudo-fraud function $f^*_\theta(\cdot)$ given $p = 2$. The blue line is the lower bound and the green line is the upper bound.

coincides with the pseudo-demand function $x^*_{\theta,1}(\cdot, \alpha_{0.25})$ associated with the first quartile $\alpha_{0.25}$ of the distribution of heterogeneity α in the population, and (standard) demand for food is strictly increasing in α (recall the discussion in Example 1.9), then the results in this application can be literally interpreted as the bounds associated with this individual household, in the event that this household chooses to commit fraud when given the incentive. In this setting, we also know that demand for fraud $f_\theta(\cdot)$ is monotonic in heterogeneity α, implying that the estimated bounds in this section imply a larger amount of fraud than 75 percent of households, and a smaller amount of fraud than 25 percent of households. With this interpretation, we can produce precise answers to the policy questions described above. Without this interpretation, it would be difficult to interpret the bounds on fraud in this section in a meaningful way, but the bounds on $b(\cdot)$ and π would still be valid because these objects are invariant across households.

1.6 Conclusion

In this chapter, I identify the necessary implications of food stamp fraud and non-parametrically estimate the extent of food stamp fraud in the United States. I find approximately \$1 billion of food stamp fraud in 2017, which is consistent with the estimate in Willey et al. (2017). I also identify and estimate several structural objects of interest—for example, (i) demand for goods (in the presence of fraud), (ii) bounds on food stamp amounts, (iii) the expected cost of an illegal exhange to the household, and (iv) bounds on demand for fraud. These objects have never been estimated before, and can be used for policy analysis.

Food stamp fraud is a concern because it redirects aid from low-income households to retailers. The

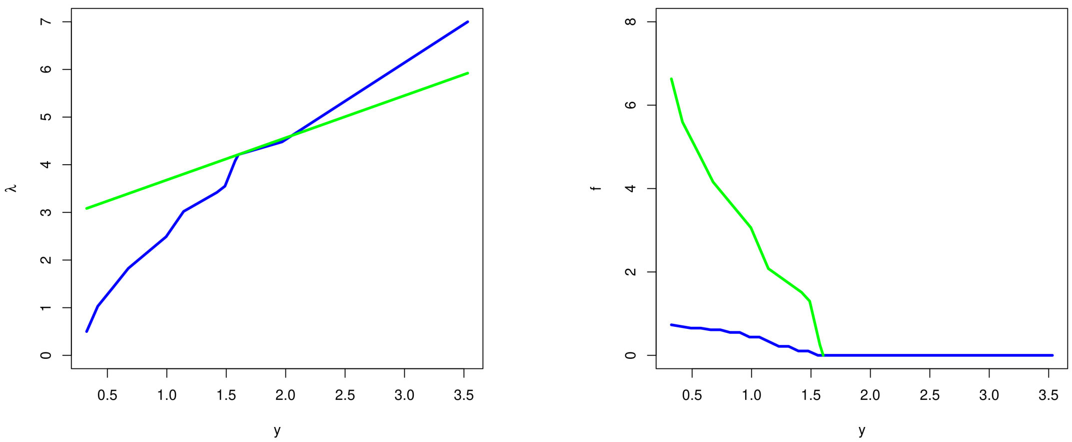

Figure 1.27. On the left, I illustrate the estimated values for $\lambda_\ell(\cdot)$ (blue) and $\lambda_h(\cdot)$ (green) given $p = 2$. On the right, I illustrate the bounds for demand for fraud $f_\theta(\cdot)$ given $p = 2$, where the blue line is the lower bound and the green line is the upper bound.

Food and Nutrition Service (FNS) is responsible for detecting and analyzing food stamp fraud, but the current methodology is expensive and limited. I construct a cheap alternative methodology that provides additional information. This chapter is the first analysis in economics that attempts to solve this critical real-world problem.

1.A Benefit Programs

The Supplemental Nutrition Assistance Program (or SNAP), formerly known as the Food Stamp Program, is a federal aid program in the United States intended to help low-income households buy food. In the 2017 fiscal year, SNAP provided \$63.6 billion in benefits to 42 million people (see Table 1.6). This appendix describes the terms and conditions of SNAP, as described in the Food Stamp Act of 1977 and the Food and Nutrition Act of 2008,[15] describes how benefit fraud is detected and analyzed in SNAP, then compares this program with the Special Supplemental Nutrition Program for Women Infants, and Children (or WIC). I focus on SNAP, rather than WIC—while these programs have similar characteristics, SNAP is a larger program with more existing information on fraud. In 2008, the Food and Nutrition Act of 2008 replaced the Food Stamp Act of 1977. For each term or condition, I cite the corresponding section in the Food and Nutrition Act of 2008. I cite the Food Stamp Act of 1977 only if there is a difference between the Food Stamp Act of 1977 and the Food and Nutrition Act of 2008 that affects the term or condition of interest.

[15] In this appendix, I do not describe the Food Stamp Act of 1964, or any of the previous versions of the Food Stamp Program. Before the Food Stamp Act of 1977, households were required to buy food stamps. This aspect was eliminated in 1979.

Table 1.6. SNAP summary by fiscal year (Food and Nutrition Service, 2018c).

Year	Average Number of Participants (in thousands)	Average Benefit Per Person (in dollars)	Total Benefits (in millions of dollars)	Total Costs (in millions of dollars)
2003	21,250	83.94	21,404.28	23,816.28
2004	23,811	86.16	24,618.89	27,099.03
2005	25,628	92.89	28,567.88	31,072.01
2006	26,549	94.75	30,187.35	32,903.06
2007	26,316	96.18	30,373.27	33,173.52
2008	28,223	102.19	34,608.40	37,639.64
2009	33,490	125.31	50,359.92	53,619.92
2010	40,302	133.79	64,702.16	68,283.47
2011	44,709	133.85	71,810.92	75,686.54
2012	46,609	133.41	74,619.34	78,411.10
2013	47,636	133.07	76,066.32	79,859.03
2014	46,664	125.01	69,998.84	74,060.33
2015	45,767	126.81	69,645.14	73,946.67
2016	44,219	125.40	66,539.35	70,912.44
2017	42,123	125.83	63,603.67	68,070.68

1.A.1 Household Definition

A "household" is defined as: (i) an individual that lives alone, (ii) an individual that lives with others but purchases and prepares food separately, or (iii) a group of individuals that live together and buy and prepare food together (Food and Nutrition Act of 2008, Section 3(m)(1)).

The following rules apply: (i) If an individual lives with her spouse, then she must be included in the same household as her spouse, (ii) if a child is under the age of 18 and living with (and under the "parental control" of) an adult, then she must be included in the same household as this adult, even if this adult is not her parent, (iii) if a child is under the age of 22 and living with a parent, then she must be included in the same household as the parent, and (iv) children cannot be assigned to more than one household—in situations with joint custody, children will be grouped with the parent that provides the most care (Food and Nutrition Act of 2008, Section 3(m)(2)).

There exist some exceptions: (i) If an individual is at least 60 years old and disabled, unable to buy and prepare food because of her disability, and the cumulative gross income of the other residents of her living situation (with the exception of her spouse) is no larger than 165 percent of the poverty guideline for a household of its size, then she is included in a household that is separate from the other residents of her living situation, with the exception of her spouse (Food and Nutrition Act of 2008, Section 3(m)(3)), (ii) if an individual is living in federally subsidized housing because she is elderly, disabled, or blind, then she is included in a household that is separate from the other residents of her living situation (Food and Nutrition Act of 2008, Section 3(m)(5)(A)), (iii) if an individual is elderly or disabled and living in a group living arrangement with less than 17 residents, then she is included in a household that is separate from the other residents of her group living arrangement (Food and Nutrition Act of 2008, Section 3(m)(5)(B)), (iv) if an individual is living in a shelter for battered women, then she is included in a household that is separate from the other residents of the shelter (Food and Nutrition Act of 2008, Section 3(m)(5)(C)), (v) if an individual is a resident of a shelter for individuals without a permanent dwelling or mailing address, then she is included in a household that is separate from the other residents of the shelter (Food and Nutrition Act of 2008, Section 3(m)(5)(D)), and (vi) if an individual is a resident of an institution intended to facilitate the recovery from a drug or alcohol addiction, then she is included in a household that is separate from the other residents of the institution, with the exception of her children (Food and Nutrition Act of 2008, Section 3(m)(5)(E)).

1.A.2 Household Eligibility

A household is *eligible* to receive benefits, only if it satisfies the criteria described in Sections 5 and 6 of the Food and Nutrition Act of 2008: (i) Every member of the household must be a citizen or lawful non-citizen of the United States (Food and Nutrition Act of 2008, Section 6(f)), (ii) every member of the household over the age of 15 and under the age of 60 must accept and maintain employment (Food and Nutrition Act of 2008, Section 6(d)(1)) unless she is (a) complying with work-registration requirements, (b) responsible for the care of a child under the age of 6, or responsible for the care of an individual that is unable to care for herself, (c) enrolled in school with at least half of a full course-load, or (d) in a program intended to facilitate the recovery from a drug or alcohol addiction (Food and Nutrition Act of 2008, Section 6(d)(2)), (iii) every member of the household that is enrolled in a school of higher education with at least half of a full course-load must be (a) under the age of 18 or over the age of 50, (b) not physically or mentally fit, (c) assigned to attend school for employment or training purposes, (d) employed for at least 20 hours per week, (e) responsible for the care of a child, or (f) receiving benefits from an admissible State program (Food and Nutrition Act of 2008, Section 6(e)), (iv) no member of the household can be fleeing to avoid prosecution or violating the conditions of a parole (Food and Nutrition Act of 2008, Section 6(k)), or convicted of any of the offenses in Section 6(r)(1) of the Food and Nutrition Act of 2008, and (v) every member of the household must cooperate with the State and provide all materials relevant for determining eligiblity (Food and Nutrition Act of 2008, Section 6(c)).

To be eligible, a household must also satisfy financial requirements that suggest that the members of the household cannot afford a nutritious diet (Food and Nutrition Act of 2008, Section 5(a)). In general, (i) household income, less admissible exclusions and deductions, must be less than the poverty guideline for a household of its size (Food and Nutrition Act of 2008, Section 5(c)(1)), (ii) if the household does not have an elderly or disabled member, then its income, less admissible exclusions, must be no larger than 130 percent of the poverty guideline for a household of the appropriate size (Food and Nutrition Act of 2008, Section 5(c)(2)), and (iii) the value of its resources—for example, cash, savings, and personal vehicles—does not exceed a threshold defined by the FNS (Food and Nutrition Act of 2008, Section 5(g)). The Food Stamp Act of 1977 defined this threshold to be $2,000 for a household without an elderly or

Table 1.7. Resource maximums in dollars (Food and Nutrition Service, 2006-2017, 2017a).

	Household Type	
Year	No Elderly/Disabled	Elderly/Disabled
1977-2010	2,000	3,000
2011-2013	2,000	3,250
2014-2016	2,250	3,250
2017	2,250	3,500

disabled member, and $3,000 for a household with an elderly or disabled member (Food Stamp Act of 1977, Section 5(g)(1)). The Food and Nutrition Act of 2008 added a condition that requires this maximum to be updated each year to account for changes in prices, as measured by the Consumer Price Index (CPI-U) in June, then rounded down to the nearest $250 increment (Food and Nutrition Act of 2008, Section 5(g)(1)(A)). Table 1.7 provides resource maximums from 2004 to 2017.

1.A.3 Poverty Guidelines

The FNS uses poverty guidelines, updated annually in the Federal Register by the U.S. Department of Health and Human Services under the authority of the Community Services Block Grant Act (Food and Nutrition Act of 2008, Section 5(c)(1)). These guidelines "do not make a distinction between farm and non-farm families, or between aged and non-aged units" (Health and Human Services Department, 2017). These guidelines are updated each year to account for changes in prices, as measured by the Consumer Price Index (CPI-U) (Health and Human Services Department, 2017). This change goes in effect, for the purpose of determining eligibility, on October 1 each year (Food and Nutrition Act of 2008, Section 5(c)). Table 1.8 provides the poverty guidelines for monthly income, in effect after October 1, for the 48 contiguous States and the District of Columbia for 2003 to 2017. There was no change in 2010 for legislative reasons, as described in Section 1012 of the Department of Defense Appropriations Act. Alaska and Hawaii have separate guidelines. Guidelines are not defined for other jurisdictions.

1.A.4 Admissible Exclusions and Deductions

According to the Food and Nutrition Act of 2008, the calculation of monthly income, for the purpose of determining eligibility, must include all forms of revenue with the exception of: (i) non-monetary benefits, (ii) unanticipated income less than or equal to $30 per quarter during the period in which benefits are received, (iii) loans, (iv) educational grants and scholarships, (v) payments, reimbursements and allowances (as described in Sections 5(d)(5), 5(d)(11) and 6(d)(I) of the Food and Nutrition Act of 2008), (vi) monetary benefits received for the care of an individual (such as child support) when this individual is not a member of the household, (vii) income earned by a member of the household under the age of 18 when this member is enrolled in elementary or secondary school, (viii) non-recurring lump-sum payments (such as tax refunds) less than or equal to $300 per quarter, (ix) income associated

Table 1.8. Poverty guidelines in dollars for the 48 contiguous States and the District of Columbia by household size and year (Food and Nutrition Service, 2018a).

	Household Size							
Year	1	2	3	4	5	6	7	8
2003	749	1,010	1,272	1,534	1,795	2,057	2,319	2,580
2004	776	1,041	1,306	1,571	1,836	2,101	2,366	2,631
2005	798	1,070	1,341	1,613	1,885	2,156	2,428	2,700
2006	817	1,100	1,384	1,667	1,950	2,234	2,517	2,800
2007	851	1,141	1,431	1,721	2,011	2,301	2,591	2,881
2008	867	1,167	1,467	1,767	2,067	2,367	2,667	2,967
2009	903	1,215	1,526	1,838	2,150	2,461	2,773	3,085
2010	903	1,215	1,526	1,838	2,150	2,461	2,773	3,085
2011	908	1,226	1,545	1,863	2,181	2,500	2,818	3,136
2012	931	1,261	1,591	1,921	2,251	2,581	2,911	3,241
2013	958	1,293	1,628	1,963	2,298	2,633	2,968	3,303
2014	973	1,311	1,650	1,988	2,326	2,664	3,003	3,341
2015	981	1,328	1,675	2,021	2,368	2,715	3,061	3,408
2016	990	1,335	1,680	2,025	2,370	2,715	3,061	3,408
2017	1,005	1,354	1,702	2,050	2,399	2,747	3,095	3,444

with changes in the cost-of-living, (x) advanced payments for earned income credits, (xi) self-support payments, (xii) admissible medical payments, and (xiii) payments for service in a combat zone (Food and Nutrition Act of 2008, Section 5(d)).

Admissible household income deductions, for the purpose of determining eligibility with respect to the requirement in the second paragraph of Appendix 1.A.2, include: (i) a *standard income deduction* defined as the maximum of (a) 8.31 percent of the poverty guideline for a household of the appropriate size, if this value is less than 8.31 percent of the poverty guideline for a household with 6 members, and 8.31 percent of the poverty guideline for a household with 6 members, otherwise, and (b) a minimum defined by the FNS (which started at \$134 in 1977, was changed to \$144 in 2008, and has been updated each year after 2008 to account for changes in the prices of commodities, as measured by the Consumer Price Index (CPI-U); see Table 1.9 for standard income deductions, in effect after October 1, for the 48 contiguous States and the District of Columbia from 2003 to 2017), (ii) 20 percent of household income, less admissible exclusions, (iii) the cost of dependent care if needed for work or education, (iv) child support payments, (v) medical expenses less than \$35 per month incurred by an elderly or disabled member of the household, and (vi) *excess shelter costs* (Food and Nutrition Act of 2008, Section 5(e); Food Stamp Act of 1977, Section 5(e)). Excess shelter costs are defined as the amount paid for housing (including, but not limited to, rent, mortgage, property taxes, and a standard utility allowance) less 50 percent of household income, after subtracting admissible exclusions and deductions (i) through (v) of the current paragraph, so long as this value is positive and below a maximum defined by the FNS (Food and Nutrition Act of 2008, Section 5(e)(6); Food and Nutrition Service, 2018a). This maximum started at \$247 in 1996. It was, then, increased four times in the period beginning in 1997, and ending in

Table 1.9. Standard income deductions in dollars for the 48 contiguous States and the District of Columbia (Food and Nutrition Service, 2018a).

	Household Size							
Year	1	2	3	4	5	6	7	8
2003	134	134	134	134	149	171	171	171
2004	134	134	134	134	153	175	175	175
2005	134	134	134	134	157	175	179	179
2006	134	134	134	134	162	186	186	186
2007	134	134	134	143	167	191	191	191
2008	144	144	144	147	172	197	197	197
2009	141	141	141	153	179	205	205	205
2010	142	142	142	153	179	205	205	205
2011	147	147	147	155	181	208	208	208
2012	149	149	149	160	187	214	214	214
2013	152	152	152	163	191	219	219	219
2014	155	155	155	165	193	221	221	221
2015	155	155	155	168	197	226	226	226
2016	155	155	155	168	197	226	226	226
2017	160	160	160	170	199	228	228	228

Table 1.10. Shelter cost maximums in dollars for the 48 contiguous States and the District of Columbia (Food and Nutrition Service, 2006-2017, 2017a).

Year	Maximum	Year	Maximum
2003	378	2011	459
2004	388	2012	469
2005	400	2013	478
2006	417	2014	490
2007	431	2015	504
2008	446	2016	517
2009	459	2017	535
2010	458		

2001, for a maximum of $340 in 2001. Since 2001, it has been updated each year to account for changes in prices, as measured by the CPI-U (Food and Nutrition Act of 2008, Section 5(e)(6)(B)). Table 1.10 provides maximums, in effect after October 1, for the 48 contiguous States and the District of Columbia from 2003 to 2017.

1.A.5 Issuance and Use of Benefits

Eligible households are given an Electronic Benefit Transfer (EBT) card (Food and Nutrition Act of 2008, Section 7(a)).[16] EBT cards can be used like debit cards at eligible retailers to purchase food approved by the U.S. Department of Agriculture under the FNS (Food and Nutrition Act of 2008, Section 7(b)). Benefits are transferred to a given household's EBT card once a month, "on or about the same date each month" (Code of Federal Regulations, Title 7, Section 274.2(d)). The date of this transfer, for a given

[16]EBT cards replaced denominated stamps in 2004 (Food and Nutrition Service, 2017b).

household, depends on a State-dependent rule for "staggering" transfers of benefits (Food and Nutrition Act of 2008, Section 7(g)(1))—for example, in Arizona, a household is transferred benefits on the first of each month if, and only if, the first letter of the last name of the applicant begins with the letter "A" or "B" (Food and Nutrition Service, 2018b). If a household does not use its benefits for an entire year, its benefits will be permanently removed from its EBT card (Food and Nutrition Act of 2008, Section 7(h)(12)(C)). If a household requires a replacement EBT card (possibly because of loss or damage), the State must provide a replacement, unless the household fails to provide an admissible explanation, as described in Section 7(h)(8)(B) of the Food and Nutrition Act of 2008. The State is allowed to charge a fee, in the form of a reduction in benefits, for replacing an EBT card (Food and Nutrition Act of 2008, Section 7(h)(8)(A)).

1.A.6 Amount of Benefits

The maximum benefit that a household can receive depends on the size of the household: This maximum is intended to encompass the cost of a "thrifty food plan" for a household of its size (Food and Nutrition Act of 2008, Sections 8(a) and 4(u)(1)), accounting for economies of scale (Food and Nutrition Act of 2008, Section 4(u)(2)). This maximum has been updated each year since 1995 to account for changes in the cost of this diet (Food and Nutrition Act of 2008, Sections 4(u)(3) and 4(u)(4)). Table 1.11 provides the maximum monthly benefits, in effect after October 1, for the 48 contiguous States and the District of Columbia from 2003 to 2017. The amount of benefits that a given household receives is defined as the maximum benefit for a household of its size less 30 percent of the household's *net income*—that is, household income less admissible exclusions and deductions, as described in Appendix 1.A.4 (Food and Nutrition Act of 2008, Section 8(a)).

1.A.7 Example

Consider a two-member household in 2017 with \$1,160 in employment income, \$230 in social security income, \$30 in medical expenses (incurred by an elderly member) and \$700 in shelter costs, each month. The household's income, less admissible exclusions, is given by the sum of employment income and social security income:

$$\$1,160 + \$230 = \$1,390. \tag{1.A.1}$$

By deducting 20 percent of employment income, the standard deduction of \$160 for a household with two members in 2017 (see Table 1.9) and medical expenses, we obtain:

$$\$1,390 - \$200 - \$160 - \$30 = \$1,000. \tag{1.A.2}$$

Table 1.11. Maximum benefit amounts in dollars for the 48 contiguous States and the District of Columbia (Food and Nutrition Service, 2018a).

	Household Size							
Year	1	2	3	4	5	6	7	8
2003	141	259	371	471	560	672	743	849
2004	149	274	393	499	592	711	786	898
2005	152	278	399	506	601	722	798	912
2006	155	284	408	518	615	738	816	932
2007	162	298	426	542	643	772	853	975
2008	176	323	463	588	698	838	926	1,058
2009	200	367	526	668	793	952	1,052	1,202
2010	200	367	526	668	793	952	1,052	1,202
2011	200	367	526	668	793	952	1,052	1,202
2012	200	367	526	668	793	952	1,052	1,202
2013	189	347	497	632	750	900	995	1,137
2014	194	357	511	649	771	925	1,022	1,169
2015	194	357	511	649	771	925	1,022	1,169
2016	194	357	511	649	771	925	1,022	1,169
2017	192	352	504	640	760	913	1,009	1,153

This amount is the household's income, less admissible exclusions and deductions, with the exception of excess shelter costs. Applying the definition in deduction (vi) of the second paragraph in Appendix 1.A.4, we obtain an excess shelter costs equal to:

$$\$700 - \frac{\$1,000}{2} = \$200. \tag{1.A.3}$$

It is worth pointing out that this value is, indeed, less than the maximum shelter cost deduction of \$535 for 2017 (see Table 1.10). The household's income, less admissible ex- clusions and deductions, is, therefore:

$$\$1,000 - \$200 = \$800. \tag{1.A.4}$$

The household satisfies financial requirement (i) in the second paragraph of Appendix 1.A.2 because the household's income, less admissible exclusions and deductions, is \$800, which is less than \$1,354, the poverty guideline for a household with two members in 2017 (see Table 1.8). The household does not need to satisfy financial requirement (ii) in the second paragraph of Appendix 1.A.2 because it has an elderly member. If the household is determined to be eligible to receive benefits with respect to the conditions in the first paragraph of Appendix 1.A.2, it will receive an amount of \$112 in benefits each month because:

$$\$352 - \$240 = \$112, \tag{1.A.5}$$

where \$352 is the maximum benefit for a two-person household in 2017 (see Table 1.11), and \$240 is 30 percent of \$800, the household's net income, computed above in (1.A.4).

1.A.8 Food and Non-Food

Households can use benefits to buy goods classified as food (Food and Nutrition Act of 2008, Section 3(k)(1))—for example, breads, cereals, fruits, vegetables, meats, and dairy products (Food and Nutrition Service, 2018d). Households can also use benefits to buy seeds or plants that can produce fruits or vegetables for consumption (Food and Nutrition Act of 2008, Section 3(k)(2)). Households cannot use benefits to buy alcohol or tobacco products (Food and Nutrition Act of 2008, Section 3(k)(1)) such as beer, wine, liquor, or cigarettes (Food and Nutrition Service, 2018d). Pet foods, soaps, paper products, household supplies, vitamins, and medicines are also excluded (Food and Nutrition Service, 2018d). Households cannot use benefits to buy "hot foods or hot food products ready for immediate consumption" (Food and Nutrition Act of 2008, Section 3(k)(1)) unless the product is (i) provided to an elderly member of the household (or her spouse) by a centre for senior citizens (Food and Nutrition Act of 2008, Section 3(k)(3)), an approved organization (Food and Nutrition Act of 2008, Section 3(k)(4)), or a living arrangement, as described in Appendix 1.A.1 (Food and Nutrition Act of 2008, Section 3(k)(7)), (ii) provided by an institution intended to facilitate the recovery from a drug or alcohol addiction (Food and Nutrition Act of 2008, Section 3(k)(5)) or a shelter for battered women (Food and Nutrition Act of 2008, Section 3(k)(8)), or (iii) provided to a member of the household without a permanent dwelling or mailing address by an approved establishment (Food and Nutrition Act of 2008, Section 3(k)(9)). In Alaska, remote households can also use benefits to buy "equipment for procuring food by hunting and fishing, such as nets, hooks, rods, harpoons, and knives (but not equipment for purposes of transportation, clothing, or shelter, and not firearms, ammunition, and explosives)" (Food and Nutrition Act of 2008, Section 3(k)(6)).

1.A.9 Eligible Retailers

Retailers must be approved by the FNS to be *eligible* to accept benefits from households as payment for food, as described in Appendix 1.A.8 (Food and Nutrition Act of 2008, Section 9(a)(1)). Approval is based on the location of the retailer, the nature of the business conducted by the retailer, the expected number of benefit transactions by the retailer if it is approved, and the reputation of the retailer (Food and Nutrition Act of 2008, Section 9(a)(1)). Prospective retailers must submit an application and submit to a visit for the purpose of determining eligibility (Food and Nutrition Act of 2008, Section 9(a)(1)). To be eligible, a retailer must (i) always carry a variety of foods, such as a grocery store, or food must account for more than half of its sales, such as a butcher (Food and Nutrition Act of 2008, Sections 3(o)(1) and 9(a)(1)), (ii) provide food, as described in conditions (i) through (iii) in Appendix 1.A.8 (Food and Nutrition Act of 2008, Sections 3(o)(1) and 9(a)(2)), or (iii) provide hunting and fishing equipment, as described at the end of Appendix 1.A.8 (Food and Nutrition Act of 2008, Sections 3(o)(1) and 9(a)(3)). If a retailer is approved, it is responsible for acquiring the necessary point-of-sale equipment

(Food and Nutrition Act of 2008, Sections 7(f)(2)).

1.A.10 Benefit Fraud Definition

The U.S. Code of Federal Regulations defines benefit fraud (also known as benefit trafficking) to mean: "(1) The buying, selling, stealing, or otherwise effecting an exchange of SNAP benefits issued and accessed via Electronic Benefit Transfer (EBT) cards, card numbers and personal identification numbers (PINs), or by manual voucher and signature, for cash or consideration other than eligible food, either directly, indirectly, in complicity or collusion with others, or acting alone; (2) The exchange of firearms, ammunition, explosives, or controlled substances, as defined in section 802 of title 21, United States Code, for SNAP benefits; (3) Purchasing a product with SNAP benefits that has a container requiring a return deposit with the intent of obtaining cash by discarding the product and returning the container for the deposit amount, intentionally discarding the product, and intentionally returning the container for the deposit amount; (4) Purchasing a product with SNAP benefits with the intent of obtaining cash or consideration other than eligible food by reselling the product, and subsequently intentionally reselling the product purchased with SNAP benefits in exchange for cash or consideration other than eligible food; or (5) Intentionally purchasing products originally purchased with SNAP benefits in exchange for cash or consideration other than eligible food. (6) Attempting to buy, sell, steal, or otherwise affect an exchange of SNAP benefits issued and accessed via Electronic Benefit Transfer (EBT) cards, card numbers and personal identification numbers (PINs), or by manual voucher and signatures, for cash or consideration other than eligible food, either directly, indirectly, in complicity or collusion with others, or acting alone" (Code of Federal Regulations, Title 7, Section 271.2).

1.A.11 Detection of Benefit Fraud

Benefit fraud is usually detected using (i) undercover investigations, (ii) social media, (iii) tips and referrals, and (iv) EBT transactions (Aussenberg, 2018). The first method involves sending undercover investigators to retailers and having these investigators attempt to exchange benefits for cash; the second method involves monitoring posts on social media to identify households attempting to exchange benefits online; the third method involves setting up and monitoring information hotlines; the fourth method involves analyzing EBT transactions to identify suspicious behaviour, indicative of benefit fraud. Unfortunately, the first method is costly, the second method provides limited information about the perpetrator, making cases difficult to pursue, and the third method is "largely unsuccessful at yielding adequate evidence of wrongdoing" (Miller et al., 2017). The fourth method is, therefore, the most sensible method for detecting benefit fraud—as a result, over 80% of (identified) benefit fraud is detected using EBT transactions (Aussenberg, 2018). In practice, transactions are analyzed with respect to a set of indicators—for example, in Tennessee, some indicators include:

(i) Transactions between 11:00PM and 5:00AM

(ii) Large even-dollar transactions, ending in ".00"

(iii) Large transactions that leave a low balance

(iv) Large transactions at small retailers

(v) Many transactions within 24 hours.

For a complete description, see Miller et al. (2017). These indicators have at least four shortcomings: (i) none of these indicators are necessary or sufficient for the existence of benefit fraud, (ii) these indicators are unable to detect certain types of benefit fraud (for example, a household may choose to exchange its benefits with a friend, or lend its EBT card to a retailer in exchange for cash, allowing this retailer to purchase food commodities at another retailer, a process known as indirect trafficking), and (iii) it is easy for a household (or a retailer) to modify its behaviour to avoid satisfying these indicators. In 2006, it was noted that, the FNS "has made good progress in its use of EBT transaction data [but it should] begin to formulate more sophisticated analyses" (Government Accountability Office, 2006). While the FNS has made progress toward this objective, this paper uses economic tools to also solve the issues described above.

1.A.12 Analysis of Benefit Fraud

The FNS is responsible for periodically analyzing the extent of benefit fraud in SNAP. In this section, I describe the methodology used by the FNS in its most recent report on benefit fraud (see Willey et al., 2017, for this report, and Aussenberg, 2018, for a broader discussion of fraud in SNAP). The objective of this report is to estimate (i) the total amount of fraud, (ii) the proportion of benefits illegally exchanged for cash, and (iii) the proportion of eligible retailers that commit fraud. The authors estimate \$1.1 billion in fraud per year, accounting for approximately 1.5 percent of all benefits, exchanged at approximately 11.8 percent of all eligible retailers. Most fraud occurred at "small" retailers, in urban areas with higher levels of poverty. According to this report, \$1.1 billion accounts for an increase of \$836 million in benefit fraud since 2005, or equivalently, a 0.5 percentage point increase in the proportion of benefits illegally exchanged for cash. These estimates are constructed using data from covert investigations into suspicious retailers, and information about retailers with suspicious EBT transactions. The authors estimate the amount of fraud by computing the proportion of investigations that found fraud, and applying a *post-stratification raking approach*. Post-stratification is used because every observation is associated with a retailer that has exhibited *suspicious behaviour*, and suspicious behaviour is thought to be highly positively correlated with *fraudulent behaviour*. Post-stratification is used to correct for the bias caused by the endogeneity of the sample. Loosely speaking, the approach adopted in this report weights observations to match the observable characteristics of the sample with the same characteristics in the population of eligible retailers. For example, if supermarkets are under-represented in the sample, then, all else equal, supermarkets will be assigned larger weights. The following characteristics were used:

(i) Size and type of the retailer

(ii) Private or public ownership of the business

(iii) Amount of poverty in the area

(iv) Urban or rural location

(v) Amount of benefits redeemed at retailers in the area.

See Appendix C in Willey et al. (2017) for a detailed description of the post-stratification procedure; see Appendix D in Willey et al. (2017) for precise definitions of the characteristics used. The authors highlight three limitations of this analysis: (i) Post-stratification can reduce bias, but "it cannot eliminate it" (Willey et al., 2017, Section 2.2). Indeed, this approach can only eliminate bias if the chosen characteristics perfectly explain the difference in the propensity to commit fraud in the sample and the population of eligible retailers. (ii) The approach is sensitive to the precise definitions of the characteristics. (iii) There is no information about the proportion of fraudulent SNAP transactions at specific retailers. The authors assume that, if a small retailer commits fraud, then 90 percent of its SNAP transactions are fraudulent, and that, if a large retailer commits fraud, then 40 percent of its SNAP transactions are fraudulent (see footnote 8 on page 6 of Willey et al., 2017). Their results rely on this assumption. There is an additional issue with the approach in this report: It is unable to detect certain types of benefit fraud. In particular, it can only detect types of benefit fraud that require a retailer—it cannot detect the other types, as described in Section 1.2.2.

1.A.13 Women, Infants, and Children (WIC)

The Special Supplemental Nutrition Program for Women, Infants, and Children (or WIC) is a federal aid program in the U.S. intended to help low-income women, infants, and children buy food. In 1972, WIC was established in an amendment to the Child Nutrition Act of 1966. Over the years, WIC has been amended several times. In this section, I describe WIC, in its current form, after the Healthy Hunger-Free Kids Act of 2010, and the major differences with SNAP.[17]

WIC is smaller than SNAP: In the 2017 fiscal year, WIC provided \$5.7 billion in benefits to 7.2 million participants (see Table 1.12). To be eligible for WIC, participants must be at nutritional risk (as established by a professional authority) and pregnant, post-partum, or breastfeeding, or under the age of five. The financial requirements for WIC are not as strict as those for SNAP—in general, income must be below 185 percent of the poverty guideline for a household of its size (see Table 1.8), and if an

[17] In this appendix, I do not describe any of the previous versions of WIC.

Table 1.12. WIC summary by fiscal year (Food and Nutrition Service, 2020d).

Year	Average Number of Participants (in thousands)	Average Benefit Per Person (in dollars)	Total Benefits (in millions of dollars)	Total Costs (in millions of dollars)
2003	7,631	35.28	3,230	4,524
2004	7,904	37.55	3,562	4,887
2005	8,023	37.42	3,602	4,992
2006	8,088	37.07	3,597	5,072
2007	8,285	39.04	3,881	5,409
2008	8,705	43.40	4,534	6,188
2009	9,122	42.40	4,640	6,471
2010	9,175	41.43	4,561	6,689
2011	8,961	46.69	5,020	7,180
2012	8,908	45.00	4,810	6,801
2013	8,663	43.26	4,497	6,501
2014	8,258	43.64	4,324	6,356
2015	8,024	43.37	4,176	6,241
2016	7,696	42.77	3,949	6,021
2017	7,286	41.24	3,606	5,705

individual is eligible for SNAP, then she automatically satisfies the income eligibility requirements for WIC. In most states, benefits are issued via EBT cards, as in SNAP, but a minority of states still use vouchers or distribute food to households directly (see Table 1.13 for a list of states that use EBT cards and statewide adoption dates, and Food and Nutrition Service, 2020c, for additional details including rollout dates). Of course, when states distribute food to households directly, benefit fraud can only exist through resale. In either case, in WIC, participants are allocated *food packages*, instead of monetary benefits, as in SNAP. Specifically, food packages consist of two forms of benefits: (i) *Prescription benefits*, which can be used to buy specific types of foods, up to maximum monthly allotments (by type), and (ii) a *cash benefit value* (CBV), that can be used, like cash, to buy fruits and vegetables. For example, a participant can be allocated a food package with a prescription that lets her buy up to 2 quarts of milk each month (in addition to prescriptions for other types of foods), and a CBV of $8 to spend on fruits and vegetables. Prescription benefits are by quantity, rather than cost. The food package that a participant receives depends on her status (e.g. infant, child, pregnant woman, etc.), not on her financial situation, as in SNAP (see Food and Nutrition Service, 2015, for a brief overview of food packages). Moreover, the list of foods that participants can buy with their prescription benefits is stricter than the list for SNAP, described in Appendix 1.A.8 (see Food and Nutrition Service, 2020a). Since maximum allotments are monthly, if benefits are not used by the end of the month, they will be lost (see, for example, Connecticut State Department of Public Health, 2020). Participants are also provided with resources, including health screening and substance-abuse services (Food and Nutrition Service, 2020b).

At the participant level, the existence of benefit fraud is, likely, much more worrisome in WIC than in SNAP: If fraud exists in WIC, then at-risk infants and children will not receive aid. That being said,

Table 1.13. Date of Statewide EBT Adoption for WIC as of April 7, 2020 (Food and Nutrition Service, 2020c).

State	EBT	Date of Adoption	State	EBT	Date of Adoption
Alabama	Yes	08/30/2019	Montana	Yes	09/22/2017
Alaska	Yes	10/17/2019	Nebraska	Yes	11/20/2018
Arizona	Yes	11/29/2017	Nevada	Yes	08/01/2009
Arkansas	Yes	07/16/2018	New Hampshire	Yes	12/28/2018
California	No	N/A	New Jersey	No	N/A
Colorado	Yes	11/01/2016	New Mexico	Yes	12/01/2007
Connecticut	Yes	06/13/2016	New York	Yes	05/31/2019
Delaware	Yes	10/24/2016	North Carolina	Yes	05/09/2018
Florida	Yes	04/01/2014	North Dakota	No	N/A
Georgia	No	N/A	Ohio	Yes	08/10/2015
Hawaii	No	N/A	Oklahoma	Yes	09/08/2016
Idaho	Yes	10/30/2019	Oregon	Yes	03/07/2016
Illinois	No	N/A	Pennsylvania	No	N/A
Indiana	Yes	09/06/2016	Rhode Island	No	N/A
Iowa	Yes	05/31/2016	South Carolina	Yes	11/18/2019
Kansas	Yes	05/30/2018	South Dakota	Yes	09/05/2017
Kentucky	Yes	10/01/2011	Tennessee	Yes	04/03/2019
Louisiana	Yes	12/23/2019	Texas	Yes	04/01/2009
Maine	No	N/A	Utah	No	N/A
Maryland	Yes	07/06/2017	Vermont	Yes	05/01/2016
Massachusetts	Yes	11/01/2014	Virginia	Yes	05/05/2014
Michigan	Yes	03/01/2009	Washington	Yes	12/31/2019
Minnesota	Yes	08/01/2019	West Virginia	Yes	10/28/2013
Mississippi	No	N/A	Wisconsin	Yes	09/23/2015
Missouri	No	N/A	Wyoming	Yes	01/01/2002

fewer households receive WIC benefits, making benefit fraud, possibly, less of a concern at the aggregate level. Some additional considerations: (i) The average amount of benefits is smaller in WIC than in SNAP (see Tables 1.6 and 1.12), possibly making fraud less likely in WIC, and implying that, if fraud exists, the amount of benefits exchanged is smaller (see the form of the fraud function $f_\theta(\cdot)$ in Corollary 1.1), (ii) since the list of goods classified as "food" is stricter in WIC than in SNAP, participants might find it harder to commit fraud in WIC, making fraud less likely, but these measures could also lead to stronger preferences for "non-food," making fraud more likely, and (iii) since participation is associated with pregnancy and childbirth, if these states-of-being lead to changes in tastes—for instance, an increase in the demand for "food" through an increase in the demand for, say, baby food—then the assumptions throughtout the paper might be violated.

1.B Proofs

1.B.1 Proof of Proposition 1.1

Consider the following proof by contradiction: Suppose $b(\cdot)$ is discontinuous at $z_0 \in \mathbb{R}^2_{++}$. Since a is continuous, z_0 satisfies $y_0 = c(p_0)$ and $b(z_0) > 0$. Since z_0 is at the threshold, increasing income from y_0 to y_1 will produce a new design $z_1 \in \mathbb{R}^2_{++}$ at which $b(z_1) = 0$. If the increase is small, so that $y_1 < y_0 + b(z_0)$, then $\psi(z_1) = y_1 < y_0 + b(z_0) = \psi(z_0)$. This implication contradicts Assumption 1.4(iii). Thus, $b(\cdot)$ is continuous in y on $\mathbb{R}_{++}$.

1.B.2 Proof of Lemma 1.1

By rearranging the inequalities in (1.2.14), we obtain:

$$\frac{x_2 - y}{\pi} \leq f \leq \frac{y + b(z) - px_1 - x_2}{1 - \pi}. \tag{1.B.1}$$

There exists an amount of fraud $f \in [0, b(z)]$ that satisfies these inequalities if, and only if, (i) the term on the left-hand side is no larger than $b(z)$, (ii) the term on the right-hand side is no smaller than zero, and (iii) the term on the left-hand side is no larger than the term on the right-hand side. All conditions are satisfied if, and only if:

$$x_2 \leq y + \pi b(z), \quad px_1 + x_2 \leq y + b(z), \quad \text{and} \quad x_2 \leq y + \big[b(z) - px_1\big]\pi. \tag{1.B.2}$$

Since $p > 0$ and $x_1 \geq 0$, the first inequality holds if the last inequality holds, and there exists $f \in [0, b(z)]$ that satisfies (1.2.14) if, and only if, both inequalities in (1.2.21) hold.

1.B.3 Proof of Lemma 1.2

The budget set $B(z, b, \pi)$ is non-empty because the household can always afford $0 \in \bar{R}$. It is compact because it is the intersection of two closed half-spaces. It is convex because its boundary is a convex function (as the minimum of two affine functions).

1.B.4 Proof of Proposition 1.3

(i) Demand $x_\theta(\cdot)$ is well-defined because utility $u(\cdot)$ is continuous and the budget set $B(z, b, \pi)$ is non-empty and compact. It is single-valued since utility $u(\cdot)$ is strictly quasi-concave (see Chapter 3.D in Mas-Colell et al., 1995, for the standard argument for linear budget sets, which can be extended to any convex budget set). It is strictly positive since $G(v)$ does not intersect the boundary of $\bar{R}$.

(ii) This result follows primarily from Assumption 1.6(ii).

1.B.5 Proof of Corollary 1.1

By budget exhaustion:

$$x_{\theta,2}(z) = \min\left\{y + b(z) - px_{\theta,1}(z),\ y + \big[b(z) - px_{\theta,1}(z)\big]\pi\right\}. \tag{1.B.3}$$

If this minimum is equal to the first term in the brackets, then:

$$px_{\theta,1}(z) + x_{\theta,2}(z) = y + b(z). \tag{1.B.4}$$

Since the right-hand side of the first inequality in (1.2.14) is strictly decreasing in the amount f, the budget constraints in (1.2.14) can only be satisfied at $f = 0$, implying $f(z) = 0$. If the minimum in (1.B.3) is equal to the second term in the brackets, then:

$$x_{\theta,2}(z) = y + \big[b(z) - px_{\theta,1}(z)\big]\pi. \tag{1.B.5}$$

Together, with the second inequality in (1.2.14), we obtain:

$$b(z) - px_{\theta,1}(z) \leq f. \tag{1.B.6}$$

Now, notice that, if this inequality is strict, then the first inequality in (1.2.14) implies:

$$b(z) - px_{\theta,1}(z) < f \leq \frac{y + b(z) - px_{\theta,1}(z) - x_{\theta,2}(z)}{1 - \pi}. \tag{1.B.7}$$

This inequality holds if, and only if:

$$x_{\theta,2}(z) < y + \big[b(z) - px_{\theta,1}(z)\big]\pi, \tag{1.B.8}$$

contradicting (1.B.5). Since the discount factor π is in $(0,1)$, the minimum in (1.B.3) is equal to the first term in the brackets if, and only if, $b(z) \geq px_{\theta,1}(z)$. Thus, we obtain:

$$f(z) = \big[b(z) - px_{\theta,1}(z)\big]^{+}. \tag{1.B.9}$$

1.B.6 Proof of Proposition 1.4

Let $B_1(z,b,\pi)$ denote the set of bundles in $\bar{R}$ that satisfy the first inequality in (1.2.21) given (z,b,π), and let $B_2(z,b,\pi)$ denote the set of bundles in $\bar{R}$ that satisfy the second inequality in (1.2.21) given (z,b,π). Each of these sets contain the budget set because:

$$B(z,b,\pi) = B_1(z,b,\pi) \cap B_2(z,b,\pi). \tag{1.B.10}$$

If $\frac{b(z)}{p} < x_{u,1}(\psi(z),p)$, then $x_u(\psi(z),p)$ maximizes $u(\cdot)$ over $B(z,b,\pi)$ since:

- $B(z,b,\pi)$ is contained in $B_1(z,b,\pi)$,
- $x_u(\psi(z),p)$ maximizes $u(\cdot)$ over $B_1(z,b,\pi)$,
- Walras' law implies $x_u(\psi(z),p)$ is in $B(z,b,\pi)$ when $\frac{b(z)}{p} < x_{u,1}(\psi(z),p)$.

Likewise, if $\frac{b(z)}{p} > x_{u,1}(\phi(z),\pi p)$, then $x_u(\phi(z),\pi p)$ maximizes $u(\cdot)$ over $B(z,b,\pi)$ since:

- $B(z,b,\pi)$ is contained in $B_2(z,b,\pi)$,
- $x_u(\phi(z),\pi p)$ maximizes u over $B_2(z,b,\pi)$,
- Walras' law implies $x_u(\phi(z),\pi p)$ is in $B(z,b,\pi)$ when $\frac{b(z)}{p} < x_{u,1}(\phi(z),\pi p)$.

It is left to show that, in every other situation, we obtain: $x_\theta(z) = (b(z)/p, y)'$. If this implication does not hold, then budget exhaustion implies (i) $px_{\theta,1}(z) > b(z)$, or (ii) $px_{\theta,1}(z) < b(z)$. In the first case, we can construct a convex combination of $x_\theta(z)$ and $x_u(\psi(z),p)$ in $B(z,b,\pi)$ (see Figure 1.21(a)); in the second case, we can construct a convex combination of $x_\theta(z)$ and $x_u(\phi(z),\pi p)$ in $B(z,b,\pi)$ (see Figure 1.21(b)). In each case, strict quasi-concavity implies that the convex combination is strictly better than demand. Consequently, we have obtained a contradiction, so that $x_\theta(z) = (b(z)/p, y)'$.

1.B.7 Proof of Proposition 1.5

Let us consider each regime in order. Since the inequalities in Proposition 1.4 define a partition of the set of $z \in \mathbb{R}^2_{++}$, it is enough to show one direction of every implication:

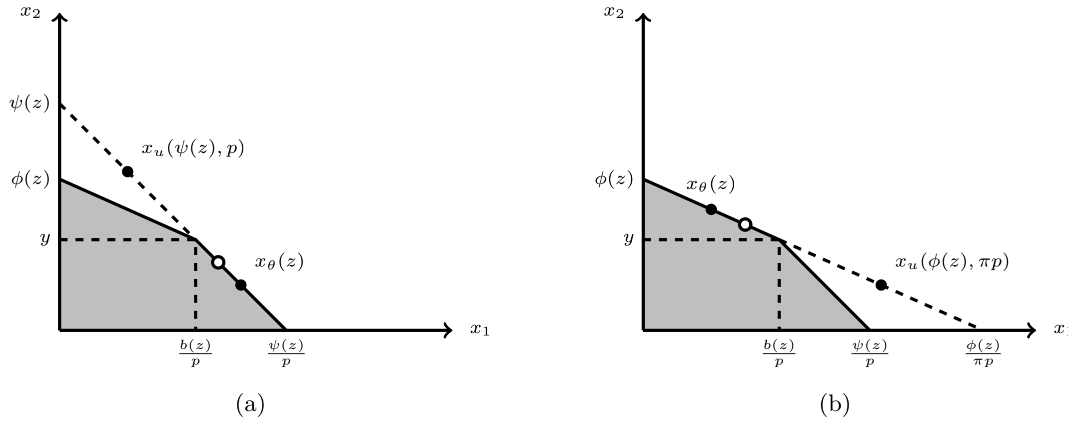

Figure 1.28. **Proof of Proposition 1.4.** On the left, there is a convex combination of $x_\theta(z)$ and $x_u(\psi(z), p)$ in $B(z, b, \pi)$; on the right, there is a convex combination of $x_\theta(z)$ and $x_u(\phi(z), \pi p)$ in $B(z, b, \pi)$. Open nodes denote examples of combinations.

(i) In the first regime:

- $b(z) < px_{u,1}(y + b(z), p) = px_{\theta,1}(z) = e_{\theta,1}(z)$.
- $e_\theta(z) = px_{\theta,1}(z) + x_{\theta,2}(z) = px_{u,1}(y + b(z), p) + x_{u,2}(y + b(z), p) = y + b(z)$.
- $e_{\theta,2}(z) = e_\theta(z) - e_{\theta,1}(z) > e_\theta(z) - b(z) = y$.

(ii) In the second regime:

- $e_{\theta,1}(z) = px_{\theta,1}(z) = b(z)$.
- $e_{\theta,2}(z) = x_{\theta,2}(z) = y$.
- $e_\theta(z) = e_{\theta,1}(z) + e_{\theta,2}(z) = y + b(z)$.

(iii) In the third regime:

- $b(z) > px_{u,1}(y + \pi b(z), \pi p) = px_{\theta,1}(z) = e_{\theta,1}(z)$.
- $y + \pi b(z) = \pi px_{u,1}(y + \pi b(z), \pi p) + x_{u,2}(y + \pi b(z), \pi p) < px_{u,1}(y + \pi b(z), \pi p) + x_{u,2}(y + \pi b(z), \pi p) = px_{\theta,1}(z) + x_{\theta,2}(z) = e_\theta(z)$.
- Budget exhaustion implies $e_\theta(z) \leq y + b(z)$. To see that this inequality is strict, first notice that, under budget exhaustion, $b(z) > px_{\theta,1}(z)$ implies $x_{\theta,2}(z) = y + \big[b(z) - px_{\theta,1}(z)\big]\pi$ (see the argument in Appendix 1.B.5). Now, in order to reach a contradiction, suppose that $e_\theta(z) = y + b(z)$, such that:

$$y + b(z) = px_{\theta,1}(z) + x_{\theta,2}(z) = px_{\theta,1}(z) + y + \big[b(z) - px_{\theta,1}(z)\big]\pi. \tag{1.B.11}$$

Rearranging this equality yields $px_{\theta,1}(z) = b(z)$, violating $b(z) > px_{\theta,1}(z)$. Since $b(z) >$

$px_{\theta,1}(z)$, budget exhaustion yields:

$$x_{\theta,2}(z) = y + \big[b(z) - px_{\theta,1}(z)\big]\pi < y. \tag{1.B.12}$$

1.B.8 Proof of Proposition 1.7

Under Assumptions 1.1 to 1.4, the policy $b(\cdot)$ is continuously-differentiable with respect to income y wherever $y \neq c(p)$. Consequently, by the relationship in Proposition 1.4 and the continuous-differentiability of standard demand $x_u(\cdot)$ on $\mathbb{R}^2_{++}$, demand x_θ is continuously-differentiable with respect to y, if it is not on the boundary of a regime and $y \neq c(p)$. For brevity, let superscript-$\circ$ denote the interior of a set. Then, we obtain:

$$\frac{\partial x_\theta(z)}{\partial y} = \begin{cases} \frac{\partial x_u(z)}{\partial y}, & \text{if } y > c(p), \\ \big(1 + \frac{\partial b(z)}{\partial y}\big) \frac{\partial x_u(y_0,p)}{\partial y_0}\big|_{y_0 = y+b(z)}, & \text{if } y < c(p) \text{ and } z \in R^\circ_{\theta,1}, \\ \big(\frac{1}{p}\frac{\partial b(z)}{\partial y}, 1\big)', & \text{if } y < c(p) \text{ and } z \in R^\circ_{\theta,2}, \\ \big(1 + \pi\frac{\partial b(z)}{\partial y}\big) \frac{\partial x_u(y_0,p)}{\partial y_0}\big|_{y_0 = y+\pi b(z)}, & \text{if } y < c(p) \text{ and } z \in R^\circ_{\theta,3}, \end{cases} \tag{1.B.13}$$

for every $z \in \mathbb{R}^2_{++}$ at which $x_\theta(\cdot)$ is not on the boundary of a regime and $y \neq c(p)$. Since $b(\cdot)$ is continuous and $\frac{\partial b(z)}{\partial y}$ is strictly larger than -1, if $b(\cdot)$ is not differentiable at $y = c(p)$, then neither is demand. Since, in the second and fourth cases in (1.B.13), the partial derivative of $x_\theta(\cdot)$ with respect to income y is strictly positive, and in the third case, it is non-positive, demand $x_\theta(\cdot)$ is not differentiable on the boundary of any of the regimes.

1.B.9 Proof of Proposition 1.8

(i) Under Assumptions 1.1 to 1.7, and N, total food income $y + b(z)$, the amount $y + \pi b(z)$, and standard demand $x_{u,1}(\cdot)$ are strictly increasing in y. Consequently:

- If $z \in R_{\theta,1} \cup R_{\theta,2}$, then total expenditure $e_\theta(\cdot)$ is strictly increasing in y since:

$$e_\theta(z) = y + b(z) = \psi(z). \tag{1.B.14}$$

- If $z \in R_{\theta,3}$, then:

$$\frac{\partial e_\theta(z)}{\partial y} = \left(1 + \pi\frac{b(z)}{\partial y}\right)\left(p\frac{\partial x_{u,1}(y_0,\pi p)}{\partial y_0} + \frac{\partial x_{u,2}(y_0,\pi p)}{\partial y_0}\right)_{y_0 = y+\pi b(z)}. \tag{1.B.15}$$

To see that this derivative is strictly positive, notice that:

$$p\frac{\partial x_{u,1}(y,\pi p)}{\partial y}+\frac{\partial x_{u,2}(y,\pi p)}{\partial y}>\pi p\frac{\partial x_{u,1}(y,\pi p)}{\partial y}+\frac{\partial x_{u,2}(y,\pi p)}{\partial y}=1. \tag{1.B.16}$$

The final equality comes from the fact that its left side is the derivative of:

$$\pi p x_{u,1}(y,\pi p)+x_{u,2}(y,\pi p)=y. \tag{1.B.17}$$

Therefore, we obtain:

$$\frac{\partial e_\theta(z)}{\partial y}>1+\pi\frac{b(z)}{\partial y}>0. \tag{1.B.18}$$

Since the policy $b(\cdot)$ and standard demand $x_u(\cdot)$ are continuous, demand $x_\theta(\cdot)$ is continuous. Consequently, the results above are sufficient for expenditure $e_\theta(\cdot)$ to be strictly increasing in income y. The result, therefore, follows from the fact that every strictly increasing real-valued function has a well-defined inverse function.

(ii) This result follows from the fact that total expenditure $e_\theta(\cdot)$ is strictly increasing.

1.B.10 Proof of Theorem 1.2

(i) This result follows from Assumptions B.1(ii) and B.1(iii), and the fact that the second and third regimes have a shared boundary (see the discussion in Remark 1.2).

(ii) First, notice that, Proposition 1.4 implies:

$$\frac{\partial e_{\theta,1}(y,p_0)}{\partial y^+}=\frac{\partial b(z_0)}{\partial y}. \tag{1.B.19}$$

Also, budget exhaustion implies that, at every z in the interior of $R_{\theta,3}$, we have:

$$\pi\frac{\partial e_{\theta,1}(z)}{\partial y}+\frac{\partial e_{\theta,2}(z)}{\partial y}=\pi p\frac{\partial x_{\theta,1}(z)}{\partial y}+\frac{\partial x_{\theta,2}(z)}{\partial y}=1+\pi\frac{\partial b(z)}{\partial y}. \tag{1.B.20}$$

Therefore, we have:

$$\pi\frac{\partial e_{\theta,1}(y,p_0)}{\partial y^-}+\frac{\partial e_{\theta,2}(y,p_0)}{\partial y^-}=1+\pi\frac{\partial b(z_0)}{\partial y}. \tag{1.B.21}$$

The form of $\delta(z_0)$ follows from rearranging this equality for the discount factor π. It is left to show that the denominator of $\delta(z_0)$ is not zero. This result follows from the fact that:

$$\frac{\partial e_{\theta,1}(y,p_0)}{\partial y^-}=p\left(1+\pi\frac{\partial b(z_0)}{\partial y}\right)\frac{\partial x_{u,1}(z_0)}{\partial y}>0\geq\frac{\partial b(z_0)}{\partial y}. \tag{1.B.22}$$

1.B.11 Proof of Theorem 1.7

Since the partial derivative of the policy $b(\cdot)$ with respect to income y is strictly larger than -1, and no larger than 0, whenever it exists, we obtain the following implications:

(i) $z \in R_{\theta,1}$ implies $\max\{0, b_\theta^*(e_h(p), p) + y_\theta^*(e_h(p), p) - y\} \leq b(z) \leq b_\theta^*(e_h(p), p)$.

(ii) $z \in R_{\theta,3}$ implies $b_\theta^*(e_\ell(p), p) \leq b(z) < b_\theta^*(e_\ell(p), p) + y_\theta^*(e_\ell(p), p) - y$.

Therefore, the result follows from the fact that:

$$e_\ell(p) = y_\theta^*(e_\ell(p), p) + b_\theta^*(e_\ell(p), p) \text{ and } e_h(p) = y_\theta^*(e_h(p), p) + b_\theta^*(e_h(p), p). \tag{1.B.23}$$

1.B.12 Proof of Lemma 1.4

First, notice that, budget exhaustion implies:

$$\pi p x_{\theta,1}^*(w) + x_{\theta,2}^*(w) = y_\theta^*(w) + \pi b_\theta^*(w), \tag{1.B.24}$$

for every $w \in W_{\theta,3}^*$. By differentiating both sides of this equality, we obtain:

$$\pi p \frac{\partial x_{\theta,1}^*(w)}{\partial e} + \frac{\partial x_{\theta,2}^*(w)}{\partial e} = \frac{\partial y_\theta^*(w)}{\partial e} \left(1 + \pi \frac{\partial b(z)}{\partial y}\right)_{y=y_\theta^*(w)}, \tag{1.B.25}$$

for every $w \in W_{\theta,3}^*$. Therefore:

$$\frac{\partial y_\theta^*(w)}{\partial e} = \left(\pi p \frac{\partial x_{\theta,1}^*(w)}{\partial e} + \frac{\partial x_{\theta,2}^*(w)}{\partial e}\right) \left(1 + \pi \frac{\partial b(z)}{\partial y}\right)_{y=y_\theta^*(w)}^{-1}, \tag{1.B.26}$$

for every $w \in W_{\theta,3}^*$. Since pseudo-demand for food $x_{\theta,1}^*(\cdot)$ is strictly increasing in total expenditure e, and the pseudo-policy $b_\theta^*(\cdot)$ is non-increasing in total expenditure, we have:

$$\frac{\partial x_{\theta,2}^*(w)}{\partial e} < \frac{\partial y_\theta^*(w)}{\partial e} < \frac{1}{1-\pi}\left(p \frac{\partial x_{\theta,1}^*(w)}{\partial e} + \frac{\partial x_{\theta,2}^*(w)}{\partial e}\right) = \frac{1}{1-\pi}, \tag{1.B.27}$$

for every $w \in W_{\theta,3}^*$.

1.B.13 Proof of Theorem 1.9

The bounds for the derivative of pseudo-income $y_\theta^*(\cdot)$ in Lemma 1.4 imply:

$$\frac{\partial x_{\theta,2}^*(w)}{\partial e} < \frac{1}{1-\pi}, \tag{1.B.28}$$

at every $w \in W^*_{\theta,3}$. Equivalently:

$$1 - \left(\frac{\partial x^*_{\theta,2}(w)}{\partial e}\right)^{-1} < \pi, \tag{1.B.29}$$

at every $w \in W^*_{\theta,3}$. Since this inequality holds at every $w \in W^*_{\theta,3}$:

$$\pi^*_0 \equiv 1 - \inf_{w \in W^*_{\theta,3}} \left(\frac{\partial x^*_{\theta,2}(w)}{\partial e}\right)^{-1} = \sup_{w \in W^*_{\theta,3}} \left\{1 - \left(\frac{\partial x^*_{\theta,2}(w)}{\partial e}\right)^{-1}\right\} \leq \pi. \tag{1.B.30}$$

1.C A Lack of Total Non-Parametric Identification

Before continuing, I present a negative result—the triple (u, π, b) is *never identified* in a completely non-parametric framework. In fact, I present a stronger result, which implies that we have to restrict the policy function $b(\cdot)$ to make inference.

Let $\mathcal{U}_0$ denote the set of all utility functions that map from $\bar{R}$ to $\mathbb{R}$ and satisfy Assumption 1.6. Let $\mathcal{B}_0$ denote the set of all policy functions that map from $\mathbb{R}^2_{++}$ to $\mathbb{R}_+$. Let $\mathcal{U}$ denote an arbitrary, non-empty subset of $\mathcal{U}_0$. Let $\mathcal{B}$ denote an arbitrary, non-empty subset of $\mathcal{B}_0$.

To be sufficiently precise, I will say that $\theta = (u, \pi, b)$ is identified on some set $\mathcal{U} \times (0,1) \times \mathcal{B}$, if θ is identified under the restriction that $\theta \in \mathcal{U} \times (0,1) \times \mathcal{B}$.

Proposition 1.9. Suppose that demand $x_\theta(\cdot)$ is observed on $\mathbb{R}^2_{++}$. Under Assumptions 1.1 to 1.7, the triple $\theta = (u, \pi, b)$ is not identified on $\mathcal{U} \times (0,1) \times \mathcal{B}_0$, for any subset of functions $\mathcal{U} \subseteq \mathcal{U}_0$.

Proof. Consider $b \in \mathcal{B}_0$ defined by: $b(z) = 0$, for all $z \in \mathbb{R}^2_{++}$. We cannot identify the discount factor π because, for every $z \in \mathbb{R}^2_{++}$, $\tilde{\pi} \in (0,1)$, and $\tilde{u} \in \mathcal{U}$, we must have:

$$x_\theta(z) = x_{\tilde{\theta}}(z), \tag{1.C.1}$$

where $\tilde{\theta} = (u, \tilde{\pi}, b)$. □

Proposition 1.9 considers the most favourable case in which demand $x_\theta(\cdot)$ is observed on the entire positive orthant, and shows that there is no identification on a set of the form $\mathcal{U} \times (0,1) \times \mathcal{B}_0$. While the proof of Proposition 1.9 is trivial, it implies that we must impose a restriction on the policy function (equivalently, on $\mathcal{B}$), if we want to identify θ. In fact, the proof of Proposition 1.9 implies that many "reasonable" restrictions on the policy function (such as continuity, boundedness, or weak monotonicity) are also insufficient.

Now, consider a non-trivial case. Let $\mathcal{B}_1$ denote the set of policies $b \in \mathcal{B}_0$ that satisfy:

$$b(y, p) \geq p'\gamma - y, \tag{1.C.2}$$

for every $z \in \mathbb{R}^2_{++}$ and some $\gamma > 0$. In words, $\mathcal{B}_1$ is the set of policy functions that ensure that the household can always purchase at least γ units of food commodities.

Proposition 1.10. Suppose that demand $x_\theta(\cdot)$ is observed on $\mathbb{R}^2_{++}$. Under Assumptions 1.1 to 1.7, the triple $\theta = (u, \pi, b)$ is not identified on $\mathcal{U}_0 \times (0,1) \times \mathcal{B}$, if the intersection $\mathcal{B} \cap \mathcal{B}_1$ is non-empty.

Proof. Consider the Stone-Geary function in Example 1. For any function, $b \in \mathcal{B}_1$, the household's demand for food commodities is bounded below by $\alpha\pi\gamma > 0$. Clearly:

$$x_\theta(z) = x_{\tilde{\theta}}(z), \tag{1.C.3}$$

for every $z \in \mathbb{R}^2_{++}$, where $\tilde{\theta} = (\tilde{u}, \pi, b)$, whenever $\tilde{u} \in \mathcal{U}_0$ coincides with $u \in \mathcal{U}_0$ on $\{x \in X : x_1 \geq \alpha\pi\gamma\}$. We can, therefore, choose any utility function that coincides with the Stone-Geary utility function on this set, whose indifference curves are "steeper" outside of this set (making non-food commodities less desirable, ensuring that the household will not consume outside of this set). $\square$

Proposition 1.10 considers the most favourable case in which demand $x_\theta(\cdot)$ is observed on the entire positive orthant, and shows that (u, π, b) is not identified on a set of the form $\mathcal{U}_0 \times (0,1) \times \mathcal{B}$, whenever $\mathcal{B}$ contains at least one element of $\mathcal{B}_1$. Since it would be unreasonable to assume away every policy function in $\mathcal{B}_1$, Proposition 1.10 implies that, in any reasonable framework, we must impose a restriction on the utility function (equivalently, on $\mathcal{U}$), if we want to identify (u, π, b). These results motivate the need for partial identification.

1.D More Bounds

In this appendix, I consider the identification of bounds for the pseudo-income function $y^*_\theta(\cdot)$ and pseudo-policy function $b^*_\theta(\cdot)$ in the first regime, when income y is not observed. Recall, $e_\ell(p)$ is the smallest amount of expenditure e in the second regime given p, and $e_h(p)$ is the largest amount of expenditure e in the second regime given p, and that, under Assumption B.2, these amounts are identified, for any observable p. By Lemma 1.3, we can identify: $y^*_\theta(e_\ell(p), p)$, $b^*_\theta(e_\ell(p), p)$, $y^*_\theta(e_h(p), p)$, and $b^*_\theta(e_h(p), p)$.

Lemma 1.8. Under Assumptions 1.1 to 1.7, N, and B.2:

$$\frac{\partial y^*_\theta(w)}{\partial e} \geq 1 \quad \text{and} \quad \frac{\partial b^*_\theta(w)}{\partial e} \leq 0, \tag{1.D.1}$$

at each $w \in W^*_{\theta,1}$.

Proof. By the *inverse function theorem*:

$$\frac{\partial y^*_\theta(w)}{\partial e} = \left(\frac{\partial e_\theta(z)}{\partial y}\right)^{-1}_{y = y^*_\theta(w)} = \left(1 + \frac{\partial b(z)}{\partial y}\right)^{-1}_{y = y^*_\theta(w)} \geq 1, \tag{1.D.2}$$

for every $w \in W^*_{\theta,1}$. By the *chain rule for differentiation*:

$$\frac{\partial b^*_\theta(w)}{\partial e} = \frac{\partial y^*_\theta(w)}{\partial e}\frac{\partial b(z)}{\partial y}\bigg|_{y=y^*_\theta(w)} \leq 0. \tag{1.D.3}$$

for every $w \in W^*_{\theta,1}$. □

Lemma 1.8 says that the rate at which the pseudo-income function $y^*_\theta(\cdot)$ increases in total expenditure e is bounded below by 1, and that the pseudo-policy function $b^*_\theta(\cdot)$ is non-increasing in total expenditure e.

Theorem 1.12. Under Assumptions 1.1 to 1.7, N, and B.2:

$$y^*_\theta(e_h(p), p) + |e_h(p) - e| \leq y^*_\theta(w) \;\text{ and }\; 0 \leq b^*_\theta(w) \leq b^*_\theta(e_h(p), p), \tag{1.D.4}$$

at each $w \in W^*_{\theta,1}$.

Proof. See Lemma 1.8. □

Theorem 1.12 implies that we can identify a lower bound for the household's income $y^*_\theta(w)$, and an upper bound for its benefits $b^*_\theta(w)$, when pseudo-demand $x^*_\theta(w)$ is in the first regime. Theorem 1.12 follows immediately from the bounds on the derivatives in Lemma 1.8. Unfortunately, we cannot say much else about these objects.

1.E Data

In this appendix, I describe the datasets used throughout this chapter. I also list the variables used in each dataset, describe how I format these specific variables for the analysis in Section 1.5, and provide a collection of relevant summary statistics.

1.E.1 Panel Survey of Income Dynamics

The Panel Survey of Income Dynamics (PSID) is a longitudinal survey in the United States. The PSID has collected information on households (and their descendants) since 1968. The original sample consists of approximately 5,000 households. About 60% of this sample is representative of the population in the United States. The remaining 40% of the sample consists of low-income households. Households are surveyed every year from 1968 to 1997, and every second year from 1999 to 2017. The PSID includes questions on household characteristics, income, benefits, and expenditure. In 1999, 2001, and 2003, households were asked whether they have been disqualified from receiving food stamps for breaking the rules since the last survey.

Variables and Formatting

Throughout this appendix, I use variables from 1999 to 2017 on (i) household size, (ii) age and sex, (iii) household income, (iv) benefit amounts, (v) region of household, (vi) grades completed, (vii) food expenditure (intended for consumption at home), (viii) total food expenditure, (ix) education, childcare, housing, transportation, and health expenditure, and (x) disqualification from the benefit program. When strictly more than one variable contains information about benefits, I choose the largest value—for example, in 1999, the household was asked for the amount of benefits that it received that year (ER14285), and in 2001, it was asked for the amount of benefits that it received two years prior (ER18371). If there is no information about benefits in a given year, then I use the amount of benefits that the household received in the previous year—for example, in 2011, the household was asked for the amount of benefits that it received in 2010 (ER48008), rather than 2011. In a similar fashion, the questions regarding household income, education expenditure, and childcare expenditure pertain to the previous year. It should also be noted that there is a debate as to whether the questions regarding food expenditure pertain to the year of the survey or the previous year (see Blundell et al., 2006, for a larger discussion). Finally, I drop all observations associated with a household income weakly below -\$100,000 or weakly above \$200,000.

Summary Statistics

I now report a collection of summary statistics for the population in the PSID. These summary statistics are intended to give the reader an overview of some basic demographics of the households in the PSID.

In Table 1.14, I report infomation on the size of the household, the age and sex of the head of the household, household income, and food stamp amounts. In this table, we see that households have 2 to 3 members, on average. We also see that the head of the household is approximately 45 years of age, on average, and that the head is male approximately 70 percent of the time. Moreover, average household income is increasing over time, from approximately \$47,000 in 1999 to approximately \$60,000 in 2017. Finally, households receive approximately \$30 to \$100 in food stamps, on average.

In Table 1.15, I report some information on the location of households. In this table, we see that approximately 15 percent of households are in the Northeast, approximately 25 percent of households are Central, approximately 45 percent of households are in the South, and approximately 15 percent of households are in the West. These approximate proportions are relatively constant from 1999 to 2017.

In Table 1.16, I report information on the highest level of education attained by the head of the household. Most heads have some college education. The second most common level of education is a high school diploma. However, approximately 20 percent of heads dropped out of high school. The least common level of education is some post-graduate education (or higher), with only 7 to 12 percent of heads reaching this level, on average.

In Table 1.17, I report the average annual expenditures of households in the PSID across several

Table 1.14. Mean household characteristics in the PSID by year. Parentheses contain standard deviations. Age and male variables correspond to the head of the household.

Year	Size	Age	Male	Income	Benefits
1999	2.78	-	0.69	47,310.13	28.58
	(1.51)		(0.45)	(36,668.89)	(154.66)
2001	2.74	-	0.73	48,010.52	38.04
	(1.48)		(0.44)	(37,629.92)	(269.66)
2003	2.69	44.93	0.69	50,901.37	44.35
	(1.47)	(16.07)	(0.46)	(38,667.27)	(223.04)
2005	2.68	44.97	0.69	52,912.61	51.47
	(1.47)	(16.38)	(0.46)	(39,852.36)	(226.02)
2007	2.65	44.87	0.68	54,881.19	56.28
	(1.46)	(16.55)	(0.46)	(41,216.69)	(274.35)
2009	2.62	45.02	0.67	57,391.52	68.89
	(1.48)	(16.64)	(0.46)	(42,286.30)	(336.38)
2011	2.58	45.04	0.66	55,050.03	94.24
	(1.47)	(16.73)	(0.47)	(42,393.69)	(426.25)
2013	2.57	45.27	0.66	56,443.29	100.63
	(1.48)	(16.78)	(0.47)	(43,583.10)	(449.57)
2015	2.53	45.47	0.65	58,048.12	80.46
	(1.48)	(16.69)	(0.47)	(44,580.42)	(362.38)
2017	2.57	45.70	0.66	60,436.24	66.70
	(1.53)	(16.61)	(0.47)	(44,541.96)	(297.36)

Table 1.15. Mean of household region indicators in the PSID by year. A household is classified as "Other" if its region is Alaska or Hawaii, a foreign country, or classified as "Wild" in 1999. A household is classified as "Central" if its region is North Central.

Year	Northeast	Central	South	West	Other
1999	0.1419	0.2547	0.4175	0.1782	0.0075
2001	0.1454	0.2555	0.4065	0.1847	0.0075
2003	0.1379	0.2497	0.4259	0.1797	0.0065
2005	0.1336	0.2506	0.4319	0.1774	0.0063
2007	0.1293	0.2534	0.4345	0.1751	0.0075
2009	0.1262	0.2476	0.4423	0.1766	0.0070
2011	0.1239	0.2467	0.4437	0.1782	0.0073
2013	0.1211	0.2479	0.4493	0.1741	0.0074
2015	0.1153	0.2492	0.4552	0.1723	0.0078
2017	0.1174	0.2468	0.4533	0.1748	0.0074

Table 1.16. Mean of indicators for the number of grades that the head of the household has completed in the PSID by year.

Year	Less than HS	HS Diploma	Some College	Some Post-Grad.
1999	0.2389	0.3223	0.3593	0.0793
2001	0.2027	0.3184	0.3915	0.0872
2003	0.2058	0.3504	0.3670	0.0767
2005	0.2029	0.3404	0.3804	0.0761
2007	0.1975	0.3377	0.3891	0.0755
2009	0.1793	0.3142	0.4074	0.0990
2011	0.1769	0.3071	0.4204	0.0955
2013	0.1705	0.2976	0.4290	0.1027
2015	0.1638	0.2926	0.4333	0.1100
2017	0.1652	0.2879	0.4328	0.1139

categories. On average, a household in the PSID spends \$5,000 to \$7,000 on food, with \$4,000 to \$5,000 of that expenditure going toward food intended for consumption at home. Moreover, on average, households approximately spend \$1,000 on education, \$450 on childcare, \$15,000 on housing, \$8,000 on transportation, and \$2,500 on health.

Demographics of Households with Positive Benefits

So far, we have seen some basic demographics of the households in the population in the PSID. However, in this paper, we are primarily interested in households that receive food stamps. I will now report the same collection of basic demographics for the subset of households receiving food stamps, and provide a brief discussion.

As before, in Table 1.18, I report information on the size of the household, the age and sex of the head of the household, household income, and food stamp amounts for households receiving a strictly positive amount of food stamps. In this table, we see that households have 3 to 4 members, on average—in other words, one additional member than the average household in the population. We also see that the head of the household is approximately 40 years of age, on average, which is slightly younger than the average household in the population. Moreover, the head is male much less often than in the population. Average household income is still increasing over time, but, as expected, average household income in this subset of households is much lower, ranging from approximately \$16,000 in 1999 to approximately \$27,000 in 2017. Last, households receive approximately \$285 to \$500 in food stamps, on average. This number is much larger than for the average household in the population, as we are conditioning on it being positive.

In Table 1.19, I report the locations of households receiving food stamps. In this table, we see that there is an increase in the proportion of households in the South, with approximately 50 to 55 percent of households located in the South (up from 40 to 45 for the population in the PSID).

In Table 1.20, I report information on the highest level of education attained by the head of the household for households receiving food stamps. It can be seen that this subset of households is much

Table 1.17. Mean household expenditures in the PSID by category and year. Parentheses contain standard deviations. The category labeled "Food" contains only food intended for consumption at home.

Year	Food	All Food	Educ.	Childcare	Housing	Trans.	Health
1999	3,763.82	5,329.51	932.97	361.91	8,699.57	7,023.21	1,597.95
	(2,558.65)	(3,387.05)	(3,344.17)	(1,374.96)	(9,746.90)	(8,790.02)	(2,737.40)
2001	4,077.17	5,899.41	1,105.31	403.36	10,024.54	8,194.98	1,818.81
	(2,600.00)	(3,549.24)	(3,948.44)	(1,397.23)	(9,198.59)	(8,855.57)	(3,259.66)
2003	4,002.45	5,788.32	994.55	378.64	9,838.56	8,194.98	2,074.75
	(2,719.08)	(3,711.70)	(3,783.73)	(1,457.44)	(8,638.06)	(9,335.36)	(4,074.95)
2005	4,238.46	6,176.76	1,084.14	394.72	14,954.16	8,646.07	2,389.18
	(3,003.52)	(4,119.48)	(4,296.68)	(1,508.63)	(20,350.82)	(9,854.31)	(3,544.56)
2007	4,419.02	6,426.07	1,207.22	464.05	16,372.44	9,008.75	2,508.46
	(3,218.58)	(4,306.02)	(4,847.03)	(1,725.55)	(16,378.64)	(9,649.63)	(3,888.88)
2009	4,467.44	6,310.79	1,068.90	490.22	16,238.66	8,312.40	2,633.52
	(3,233.31)	(4,300.32)	(4,191.54)	(1,943.52)	(15,332.47)	(9,171.05)	(5,077.66)
2011	4,611.03	6,490.52	1,137.13	476.94	16,206.90	8,529.50	2,495.61
	(3,401.04)	(4,493.06)	(4,795.37)	(1,897.57)	(23,584.73)	(8,844.75)	(4,331.85)
2013	4,736.18	6,710.32	1,167.34	513.92	15,485.16	8,549.24	3,030.81
	(3,550.09)	(4,736.60)	(5,005.36)	(2,030.45)	(11,331.18)	(7,186.32)	(5,494.64)
2015	4,997.19	7,125.71	1,080.77	466.69	15,741.27	8,039.64	3,184.14
	(3,629.32)	(4,853.50)	(4,868.01)	(1,880.42)	(10,772.16)	(7,186.32)	(5,376.42)
2017	5,513.09	7,926.83	1,028.12	526.33	16,478.80	8,087.99	3,064.33
	(4,023.64)	(5,345.61)	(4,957.76)	(2,083.30)	(11,397.28)	(7,570.24)	(4,519.89)

Table 1.18. Mean household characteristics for households with positive benefits in the PSID by year. Parentheses contain standard deviations. Age and male variables correspond to the head of the household.

Year	Size	Age	Male	Income	Benefits
1999	3.50	-	0.35	16,694.36	285.60
	(1.84)		(0.47)	(14,919.02)	(407.16)
2001	3.41	-	0.42	15,849.16	458.57
	(1.73)		(0.49)	(14,231.05)	(827.46)
2003	3.33	39.85	0.38	19,185.19	387.69
	(1.75)	(14.89)	(0.48)	(16,630.81)	(549.51)
2005	3.26	38.98	0.43	22,445.28	363.75
	(1.73)	(14.30)	(0.49)	(20,264.84)	(497.59)
2007	3.25	39.31	0.41	21,365.56	383.63
	(1.74)	(14.58)	(0.49)	(18,443.39)	(622.70)
2009	3.16	39.16	0.45	24,573.80	428.05
	(1.74)	(14.11)	(0.49)	(21,640.80)	(741.39)
2011	3.09	39.18	0.44	23,689.59	499.19
	(1.74)	(14.40)	(0.49)	(21,878.68)	(872.09)
2013	3.08	39.96	0.45	24,330.61	472.70
	(1.71)	(14.60)	(0.49)	(21,610.73)	(879.67)
2015	2.96	41.11	0.43	24,593.77	412.67
	(1.70)	(14.82)	(0.49)	(22,275.78)	(732.58)
2017	3.09	42.36	0.45	27,807.92	369.42
	(1.83)	(15.13)	(0.49)	(23,905.59)	(614.87)

Table 1.19. Mean of household region indicators for households with positive benefits in the PSID by year. A household is classified as "Other" if its region is Alaska or Hawaii, a foreign country, or classified as "Wild" in 1999. A household is classified as "Central" if its region is North Central.

Year	Northeast	Central	South	West	Other
1999	0.0843	0.2398	0.5058	0.1700	0.0000
2001	0.0954	0.2234	0.5292	0.1496	0.0021
2003	0.0744	0.2737	0.5189	0.1317	0.0011
2005	0.0846	0.2586	0.5428	0.1138	0.0000
2007	0.0863	0.2615	0.5299	0.1205	0.0017
2009	0.0868	0.2649	0.5264	0.1218	0.0000
2011	0.0885	0.2699	0.5139	0.1275	0.0000
2013	0.0933	0.2604	0.5306	0.1150	0.0005
2015	0.0829	0.2619	0.5310	0.1229	0.0011
2017	0.1037	0.2539	0.5201	0.1221	0.0000

Table 1.20. Mean of indicators for the number of grades that the head of the household has completed for households with positive benefits in the PSID by year.

Year	Less than HS	HS Diploma	Some College	Some Post-Grad.
1999	0.5327	0.2943	0.1682	0.0046
2001	0.5373	0.2733	0.1799	0.0093
2003	0.4538	0.3341	0.2071	0.0047
2005	0.4390	0.3257	0.2266	0.0085
2007	0.4133	0.3562	0.2232	0.0071
2009	0.3742	0.3422	0.2713	0.0121
2011	0.3446	0.3446	0.3005	0.0101
2013	0.3211	0.3421	0.3211	0.0154
2015	0.3351	0.3357	0.3107	0.0183
2017	0.3193	0.3437	0.3149	0.0219

less educated than the population. In particular, most heads in this subset drop out of high school, and the second most common level of education is a high school diploma, followed by some college education. Approximately 1 percent of the heads in this subset complete some post-graduate education, which is down from approximately 7 to 12 percent.

In Table 1.21, I report the average annual expenditures of households receiving food stamps across several categories. On average, a household receiving food stamps spends \$3,000 to \$4,000 on food, with \$2,000 to \$3,000 of that expenditure going toward food intended for consumption at home. Moreover, on average, these households approximately spend \$300 on education, \$300 on childcare, \$8,000 on housing, \$4,000 on transportation, and \$800 on health. We see a big decrease in expenditures across all categories because we are now looking at a subset of households with low income.

Table 1.21. Mean household expenditures for households with positive benefits in the PSID by category and year. Parentheses contain standard deviations. The category labeled "Food" contains only food intended for consumption at home.

Year	Food	All Food	Educ.	Childcare	Housing	Trans.	Health
1999	2,107.32	2,787.23	172.15	263.92	5,630.29	2,608.38	529.45
	(2,085.83)	(2,487.86)	(583.15)	(1,064.20)	(16,182.20)	(4,030.75)	(1,304.12)
2001	2,264.90	3,112.36	309.11	284.75	5,630.29	3,670.69	650.98
	(2,287.08)	(2,732.06)	(1,592.18)	(818.17)	(3,671.39)	(5,237.90)	(1,700.39)
2003	2,229.05	2,987.03	297.61	348.16	5,718.27	3,824.09	646.82
	(2,344.31)	(2,814.64)	(1,650.09)	(1,358.52)	(3,984.87)	(5,220.45)	(1,670.26)
2005	2,303.16	3,275.34	323.24	247.39	8,826.48	4,299.16	814.18
	(2,513.51)	(3,204.87)	(1,419.35)	(1,019.18)	(6,493.59)	(5,407.22)	(1,760.71)
2007	2,459.90	3,409.46	318.90	377.74	9,795.78	4,582.17	858.36
	(2,801.13)	(3,350.66)	(1,342.04)	(1,432.66)	(10,382.97)	(6,014.05)	(3,470.12)
2009	2,496.34	3,459.11	346.71	385.74	9,888.89	4,577.50	863.62
	(2,831.42)	(3,398.60)	(1,846.56)	(1,584.36)	(6,599.71)	(5,786.23)	(2,435.85)
2011	2,455.53	3,421.13	478.40	322.05	10,457.84	5,446.29	813.60
	(2,964.05)	(3,521.90)	(2,403.32)	(1,114.12)	(7,276.06)	(6,487.76)	(2,755.98)
2013	2,529.35	3,621.55	566.46	437.08	10,510.73	5,380.46	999.37
	(2,729.82)	(3,927.60)	(3,066.41)	(1,729.64)	(7,303.56)	(6,387.22)	(5,709.01)
2015	2,753.60	3,857.22	452.28	248.23	10,651.76	4,774.37	1,165.99
	(2,911.82)	(3,653.36)	(2,437.40)	(1,008.94)	(6,713.30)	(5,296.86)	(6,682.28)
2017	3,240.87	4,633.66	349.17	335.35	11,862.89	4,829.04	869.08
	(3,425.63)	(4,275.49)	(1,907.78)	(1301.68)	(6,620.83)	(5,234.05)	(2,101.80)

Demographics of Disqualified Households

I now provide the statistics above for the subset of households that have been disqualified from receiving food stamps for breaking the rules. In Section 1.5, we saw the number and proportion of these households in the years in which this question was asked. I now focus only on the demographics of these households.

In Table 1.22, I report information on the size of the household, the age and sex of the head of the household, and household income for households that have been disqualified from receiving food stamps for breaking the rules. I do not report information about benefits because it is difficult to interpret this information (as these households have been disqualified). Once again, we see that disqualified households are larger, have a younger head, are more likely to have a female head, and have lower income than the population in the PSID. That being said, the demographics of these disqualified households appear to be quite similar to the demographics of the subset of households receiving a positive amount of food stamps. While these numbers are calculated using a relatively small number of households, these numbers suggest that it is possible that, on average, disqualified households have an even younger head, and are even more likely to have a female head.

In Table 1.23, I report the locations of the households that have been disqualified from receiving food stamps for breaking the rules. In this table, we once again see that there is a larger proportion of households in the South. This proportion is larger than both the proportion in the population as a whole, and the proportion in the subset of households receiving food stamps, with approximately 60 to

Table 1.22. Mean household characteristics for disqualified households in the PSID by year. Parentheses contain standard deviations. Age and male variables correspond to the head of the household.

Year	Size	Age	Male	Income
1999	3.50	-	0.45	17,114.27
	(1.92)		(0.50)	(12,654.93)
2001	3.14	-	0.28	13,600.00
	(1.77)		(0.48)	(8,521.54)
2003	3.46	35.93	0.26	16,434.20
	(1.76)	(8.88)	(0.45)	(11,323.18)

Table 1.23. Mean of household region indicators for disqualified households in the PSID by year. A household is classified as "Other" if its region is Alaska or Hawaii, a foreign country, or classified as "Wild" in 1999. A household is classified as "Central" if its region is North Central.

Year	Northeast	Central	South	West	Other
1999	0.0909	0.0909	0.5909	0.2272	0.0000
2001	0.2857	0.1428	0.5714	0.0000	0.0000
2003	0.0666	0.2000	0.6666	0.0666	0.0000

65 percent of households located in the South (up from 40 to 45 for the population in the PSID and 50 to 55 for the subset of households receiving food stamps). This number suggests that either households in the South are more likely to break the rules, or more likely to get caught, or both.

In Table 1.24, I report information on the highest level of education attained by the head of the household for the households that have been disqualified from receiving food stamps for breaking the rules. It can be seen that disqualified households are mostly less educated than the population, but more educated than the subset of households receiving food stamps. It is natural to imagine that breaking the rules, which requires an understanding of the food stamp system and how to manipulate it, is associated with a higher level of education.

In Table 1.25, I report the average annual expenditures of disqualified households across several categories. On average, a disqualified household spends \$3,000 to \$4,500 on food, with \$2,000 to \$3,500 of that expenditure going toward food intended for consumption at home. Moreover, on average, these households approximately spend \$600 on childcare, \$6,000 on housing, \$2,500 on transportation, and \$200 on health. Spending on education varies widely from \$100 to approximately \$2,500, on average, depending on the year. Once again, we see a big decrease in expenditures across most categories. Most

Table 1.24. Mean of indicators for the number of grades that the head of the household has completed for disqualified households in the PSID by year.

Year	Less than HS	HS Diploma	Some College	Some Post-Grad.
1999	0.4761	0.3809	0.1428	0.0000
2001	0.2857	0.0000	0.7142	0.0000
2003	0.3333	0.4666	0.2000	0.0000

Table 1.25. Mean household expenditures for disqualified households in the PSID by category and year. Parentheses contain standard deviations. The category labeled "Food" contains only food intended for consumption at home.

Year	Food	All Food	Educ.	Childcare	Housing	Trans.	Health
1999	2,830.00	3,638.86	122.27	492.77	5,679.22	1,345.00	220.18
	(2,187.84)	(2,612.29)	(326.77)	(1,992.41)	(3,847.82)	(1,736.90)	(427.67)
2001	3,502.28	4,765.14	100.00	802.85	6,261.71	2,566.57	189.28
	(2,541.94)	(2,817.09)	(191.48)	(1,269.14)	(3,141.39)	(3,079.83)	(318.48)
2003	2,098.66	3,148.00	2,501.00	409.06	6,167.53	3,541.60	381.13
	(1,516.52)	(2,268.96)	(9,132.74)	(820.76)	(3,910.71)	(3,731.22)	(781.33)

of these numbers are even below those for the population of households receiving food stamps. Since these households report an annual income that is similar to the population of households receiving food stamps, this result suggests that these households are spending their money in categories that are not in this table. There is one major exception: These households appear to spend more on childcare. It is possible that this result is an implication of random variation, but it is also possible that households with higher childcare costs are financially constrained, and, as a result, more likely to break the rules.

Demographics of Households Spending Less on Food than Benefits

I now provide the statistics above for the subset of households that report spending less on food (intended for consumption at home) than they receive in food stamps. As described throughout the paper, this expenditure pattern holds if, and only if, the household is committing fraud. While we could observe households with this expenditure pattern due to measurement error, it is worth investigating the demographics of these households, and how they differ from the demographics described above.

In Table 1.26, I report information on the size of the household, the age and sex of the head of the household, household income, and benefits for households that report spending less on food (intended for consumption at home) than they receive in food stamps. Similar to disqualified households, we see that the demographics of these households are similar to the subset of households receiving a positive amount of food stamps, but the household head is younger, on average, and more likely to be female. In addition to these differences, we now also see lower income and higher benefits, on average. This final result is natural because we are conditioning on households that report spending less on food (intended for consumption at home) than they receive in food stamps, and, all else equal, poorer households receive more in benefits.

In Table 1.27, I consider the locations of the households that report spending less on food (intended for consumption at home) than they receive in food stamps. Once again, we see a larger proportion of households in the South, when we compare with the population or the subset of households that receive a positive amount of benefits, although the difference is slightly smaller than for disqualified households.

In Table 1.28, I report information on the highest level of education attained by the head of the

Table 1.26. Mean household characteristics for households spending less of food than receiving in benefits in the PSID by year. Parentheses contain standard deviations. Age and male variables correspond to the head of the household.

Year	Size	Age	Male	Income	Benefits
1999	3.45	-	0.28	12,062.66	401.28
	(1.86)		(0.45)	(10,007.62)	(642.28)
2001	3.38	-	0.37	11,945.78	765.07
	(1.89)		(0.48)	(11,409.07)	(1,294.75)
2003	3.31	36.47	0.28	14,296.85	617.91
	(1.75)	(13.86)	(0.45)	(11,674.89)	(837.57)
2005	3.09	36.60	0.34	15,263.69	502.83
	(1.70)	(13.36)	(0.47)	(15,159.82)	(720.54)
2007	3.17	37.66	0.35	16,243.33	567.57
	(1.77)	(14.99)	(0.47)	(14,540.90)	(1,004.94)
2009	3.12	36.22	0.37	18,679.29	643.12
	(1.68)	(12.80)	(0.48)	(17,655.23)	(1,178.75)
2011	3.02	37.57	0.37	16,433.13	765.06
	(1.77)	(14.06)	(0.48)	(16,008.28)	(1,343.25)
2013	3.06	38.04	0.37	17,401.52	796.85
	(1.67)	(14.22)	(0.48)	(17,260.75)	(1,473.27)
2015	2.90	38.92	0.36	15,939.15	679.66
	(1.83)	(14.35)	(0.48)	(14,448.53)	(1,248.10)
2017	2.97	41.17	0.37	18,223.85	529.58
	(1.86)	(15.19)	(0.48)	(18,398.51)	(1,022.28)

Table 1.27. Mean of household region indicators for households spending less of food than receiving in benefits in the PSID by year. A household is classified as "Other" if its region is Alaska or Hawaii, a foreign country, or classified as "Wild" in 1999. A household is classified as "Central" if its region is North Central.

Year	Northeast	Central	South	West	Other
1999	0.0580	0.2903	0.5354	0.1161	0.0000
2001	0.0782	0.1913	0.5391	0.1913	0.0000
2003	0.0913	0.2608	0.5652	0.0826	0.0000
2005	0.0873	0.2718	0.5372	0.1035	0.0000
2007	0.0654	0.2827	0.5476	0.1041	0.0000
2009	0.0691	0.2939	0.5360	0.1008	0.0000
2011	0.0726	0.2960	0.5344	0.0968	0.0000
2013	0.0894	0.2783	0.5328	0.0994	0.0000
2015	0.0829	0.2890	0.5284	0.0995	0.0000
2017	0.1046	0.2617	0.5344	0.0991	0.0000

Table 1.28. Mean of indicators for the number of grades that the head of the household has completed for households spending less of food than receiving in benefits in the PSID by year.

Year	Less than HS	HS Diploma	Some College	Some Post-Grad.
1999	0.4930	0.3194	0.1875	0.0000
2001	0.5523	0.2285	0.2095	0.0095
2003	0.4955	0.3318	0.1681	0.0044
2005	0.4256	0.3682	0.1993	0.0067
2007	0.4223	0.3478	0.2298	0.0000
2009	0.4058	0.3470	0.2352	0.0117
2011	0.3802	0.3593	0.2528	0.0076
2013	0.3939	0.3535	0.2383	0.0141
2015	0.3951	0.3390	0.2512	0.0146
2017	0.3464	0.3746	0.2676	0.0112

household for households that report spending less on food (intended for consumption at home) than they receive in food stamps. Unlike disqualified households, these proportions are very similar to the proportions for the population of households receiving food stamps, but now fewer households have some college education.

In Table 1.29, I report the average annual expenditures of households that report spending less on food (intended for consumption at home) than they receive in food stamps across several categories. On average, these households spend \$550 to \$1,000 on food, with \$50 to \$150 of that expenditure going toward food intended for consumption at home. These numbers are extremely small. This result is consistent with the fact that benefits cannot exceed a maximum allotment, and the fact that we are conditioning on food expenditure being below the benefit allotment. Moreover, on average, these households approximately spend \$300 on education, \$200 on childcare, \$6,000 on housing, \$3,000 on transportation, and \$500 on health.

1.E.2 Nielsen Homescan Consumer Panel

The Nielsen Homescan Consumer Panel (NHCP) is a longitudinal dataset in the United States. The NHCP tracks the purchases of households. It started in 2004 and follows approximately 40,000 to 60,000 households each year. Some households participate for multiple years, while others only participate for a short amount of time. The sample is intended to be representative of the population in the United States. The NHCP includes questions on household characteristics and income.

Nielsen provides each household with a barcode scanner. Households scan all purchased goods. Prices are entered by the household or linked with retailer data. Households are financially compensated with benefits and lotteries, and self-select into participation. This final feature could create a self-selection bias that is neglected throughout this chapter.

Table 1.29. Mean household expenditures for households spending less of food than receiving in benefits in the PSID by category and year. Parentheses contain standard deviations. The category labeled "Food" contains only food intended for consumption at home.

Year	Food	All Food	Educ.	Childcare	Housing	Trans.	Health
1999	49.09	556.12	154.66	187.09	4,580.96	2,191.26	369.51
	(262.25)	(811.13)	(625.25)	(553.67)	(2,744.09)	(3,468.47)	(1,056.06)
2001	150.43	678.95	306.13	166.91	4,542.88	2,409.58	411.98
	(519.29)	(960.74)	(1,067.54)	(577.55)	(3,186.82)	(3,921.06)	(1,790.23)
2003	110.43	589.53	247.53	245.42	4,668.50	2,416.73	617.59
	(457.69)	(851.57)	(1,228.82)	(903.77)	(3,403.38)	(3,383.91)	(2,236.96)
2005	77.83	669.47	275.47	220.34	7,030.40	2,680.96	454.97
	(393.67)	(919.81)	(1,833.60)	(756.59)	(5,320.78)	(3,773.57)	(1,189.99)
2007	114.30	811.60	217.02	428.17	7,528.86	3,274.60	533.07
	(596.87)	(1,382.04)	(953.63)	(1,520.27)	(8,430.94)	(4,537.06)	(1,433.95)
2009	75.67	738.99	167.97	445.96	8,037.27	3,410.27	498.35
	(371.34)	(1,092.68)	(700.82)	(2,050.84)	(6,097.02)	(5,157.61)	(2,440.51)
2011	143.57	771.04	295.38	231.91	8,379.49	3,894.10	452.04
	(575.15)	(1,132.14)	(1,262.92)	(919.29)	(5,960.74)	(5,910.49)	(1,470.67)
2013	139.85	866.13	473.39	455.04	8,511.40	3,625.74	445.42
	(564.79)	(1,398.20)	(3,368.63)	(2,136.80)	(5,537.70)	(5,289.77)	(1,222.02)
2015	114.58	803.57	385.56	204.63	7,904.00	3,210.60	271.04
	(511.40)	(1,215.51)	(2,747.87)	(931.46)	(5,488.67)	(4,836.76)	(809.12)
2017	94.93	994.41	242.78	279.21	9,803.57	2,890.06	341.99
	(451.35)	(1,369.99)	(1,532.73)	(1,350.38)	(4,821.16)	(3,649.30)	(1,076.24)

Variables and Formatting

In the NHCP, goods are grouped into eleven distinct departments: (i) health and beauty aids, (ii) dry grocery, (iii) frozen foods, (iv) dairy, (v) deli, (vi) packaged meat, (vii) fresh produce, (viii) non-food grocery, (ix) alcohol, (x) general merchandise, and (xi) "magnet data" products (consisting of products without Universal Product Codes such as fresh produce). I group health and beauty aids with general merchandise, fresh produce with magnet data products, and deli with packaged meat, for a total of eight categories of homogeneous goods.

For each household, I sum the total amount paid (less coupon values) for every transaction within each department and month. I restrict attention to August to October in 2016,[18] and keep all households with positive expenditure in all categories in each month. This procedure leaves us with three observations for 4,807 households for a total of 14,421 observations. The price of a specific category faced by a household in a month is calculated by dividing the household's total expenditure in that category and month by the quantity purchased by the household in that category and month.

I create a benchmark bundle, as described in Section 1.4.1, by taking the sum of all expenditures within each category in August and dividing it by the sum of all prices within that category in August. I classify dry grocery, frozen foods, dairy, deli and packaged meat, and fresh produce and magnet data

[18]As mentioned in Section 1.5.2, these three months are consecutive, and avoid holidays (on which consumption might be irregular) such as Independence Day, Christmas Day, and New Year's Eve. The short time frame reduces the possibility of changing tastes or product availability.

Table 1.30. Household size in the NHCP sample and in the 2017 Annual Social and Economic Supplement (ASEC) of the CPS. CPS numbers are in thousands.

	Sample		CPS	
Size	Number	Proportion	Number	Proportion
1	665	0.1383	35,388	0.2812
2	2,644	0.5500	42,785	0.3400
3	700	0.1456	19,423	0.1543
4	537	0.1117	16,267	0.1292
5	177	0.0368	7,548	0.0599
6	58	0.0120	2,813	0.0223
7+	16	0.0054	1,596	0.0126
Total	4,797	1.0000	125,819	1.0000

products as “food” and all other categories as “non-food.” I construct the price of food p_{i1t} faced by household i in month t by calculating the cost of the food products in the benchmark bundle for that household in that month. I construct the price of non-food p_{i2t} in a similar fashion. The normalized price of food p_{it} is calculated by taking the ratio of these prices such that $p_{it} = p_{i1t}/p_{i2t}$. I construct the quantity of food x_{i1t} bought by household i in month t by calculating the cost of the food products in household i's chosen bundle in month t, and dividing this cost by the price of food p_{i1t}. I construct the quantity of non-food x_{i2t} in a similar fashion. After constructing $p_{it}$and x_{it}, I construct normalized expenditure e_{it} using the relationship: $p_{it}x_{i1t} + x_{i2t} = e_{it}$.

Summary Statistics

I now report summary statistics and discuss the representativeness of the NHCP sample (after the formatting above) by comparing these statistics with those found in the Current Population Survey (CPS).

In Table 1.30, we see the distribution of household size in the NHCP and the CPS. These distributions are similar, but the NHCP has a smaller proportion of households with a single member, and a larger proportion of households with exactly two members. Likely, single-member households do not meet the expenditure requirements described above as often as larger households.

In Table 1.31, we see the distribution of household income in the NHCP and the CPS. Once again, these distributions are similar. The NHCP has a higher proportion of households with annual income between \$70,000 and \$99,999. It is possible that this difference is due to the difference in the distributions of household size.

In Table 1.32, we see the distribution of the age of the eldest head of the household in the NHCP and the age of the householder in the CPS. The NHCP has older heads, but this difference can be explained by the fact that, for the NHCP, I am reporting the age of the eldest head, and for the CPS, I am reporting the age of the householder, which might be younger than the eldest head.

Table 1.31. Annual household income in the NHCP sample and in the 2017 Annual Social and Economic Supplement (ASEC) of the CPS. CPS numbers are in thousands.

	Sample		CPS	
Income	Number	Proportion	Number	Proportion
Under $5,000	24	0.0049	4,138	0.0327
$5,000 to $9,999	34	0.0070	3,878	0.0307
$10,000 to $14,999	95	0.0197	6,122	0.0485
$15,000 to $19,999	119	0.0247	5,838	0.0462
$20,000 to $24,999	176	0.0366	6,245	0.0494
$25,000 to $29,999	212	0.0441	5,939	0.0470
$30,000 to $34,999	245	0.0509	5,919	0.0468
$35,000 to $39,999	238	0.0495	5,727	0.0453
$40,000 to $44,999	224	0.0465	5,487	0.0434
$45,000 to $49,999	276	0.0574	5,089	0.0403
$50,000 to $59,999	519	0.1079	9,417	0.0746
$60,000 to $69,999	436	0.0907	8,213	0.0650
$70,000 to $99,999	1,172	0.2438	19,249	0.1524
$100,000+	1,037	0.2157	34,963	0.2769
Total	4,807	1.0000	126,224	1.0000

Table 1.32. Age of eldest household head in the NHCP sample and the householder in the 2017 Annual Social and Economic Supplement (ASEC) of the CPS. CPS numbers are in thousands.

	Sample		CPS	
Age	Number	Proportion	Number	Proportion
Under 25	6	0.0012	6,361	0.0505
25 to 29	46	0.0095	9,453	0.0751
30 to 34	150	0.0312	10,594	0.0842
35 to 39	241	0.0501	10,651	0.0846
40 to 44	270	0.0561	10,571	0.0840
45 to 49	366	0.0761	11,115	0.0883
50 to 54	557	0.1158	12,180	0.0968
55 to 64	1,477	0.3072	23,896	0.1899
65+	1,694	0.3524	30,999	0.2463
Total	4,807	1.0000	125,819	1.0000

1.F More on Conditional Quantile Estimation

In this appendix, I describe the details that are needed to obtain the estimates of the conditional quantiles in Figures 1.22 and 1.23, and provide estimates at other quantiles for each of the two samples considered.

First, I will briefly describe the details of the construction of the subset of the NHCP consisting of poorer households with more members and younger female heads. This sample consists of all observations associated with households that have (i) two or more members, (ii) household income less than $25,000, and (iii) a female head that is no older than 40 years of age. After conditioning on these demographics, we are left 3 observations for 248 households for a total of 744 observations.

In the full NHCP sample, I estimate the conditional quantiles of food consumption using $\lambda_0 = \lambda_1 = 2$ and $\lambda_2 = 5$ as tuning parameters. In the subset of the NHCP, I use $\lambda_0 = \lambda_1 = 0.7$ and $\lambda_2 = 1.5$ as tuning parameters when $\tau = 0.1$ and $\tau = 0.25$, and then increase λ_0 to 1 and λ_2 to 1.9 for all other conditional quantiles.

In Figure 1.29, I plot the estimated conditional quantiles of food consumption for $\tau = 0.10, 0.75, 0.90$. In each row, on the left, I plot the estimate for the full NHCP sample, and on the right, I plot the estimate for the subset of the NHCP containing poorer households with more members and younger female heads. In this figure, we see that there is no prominent ridge or valley in either sample at any of these quantiles.

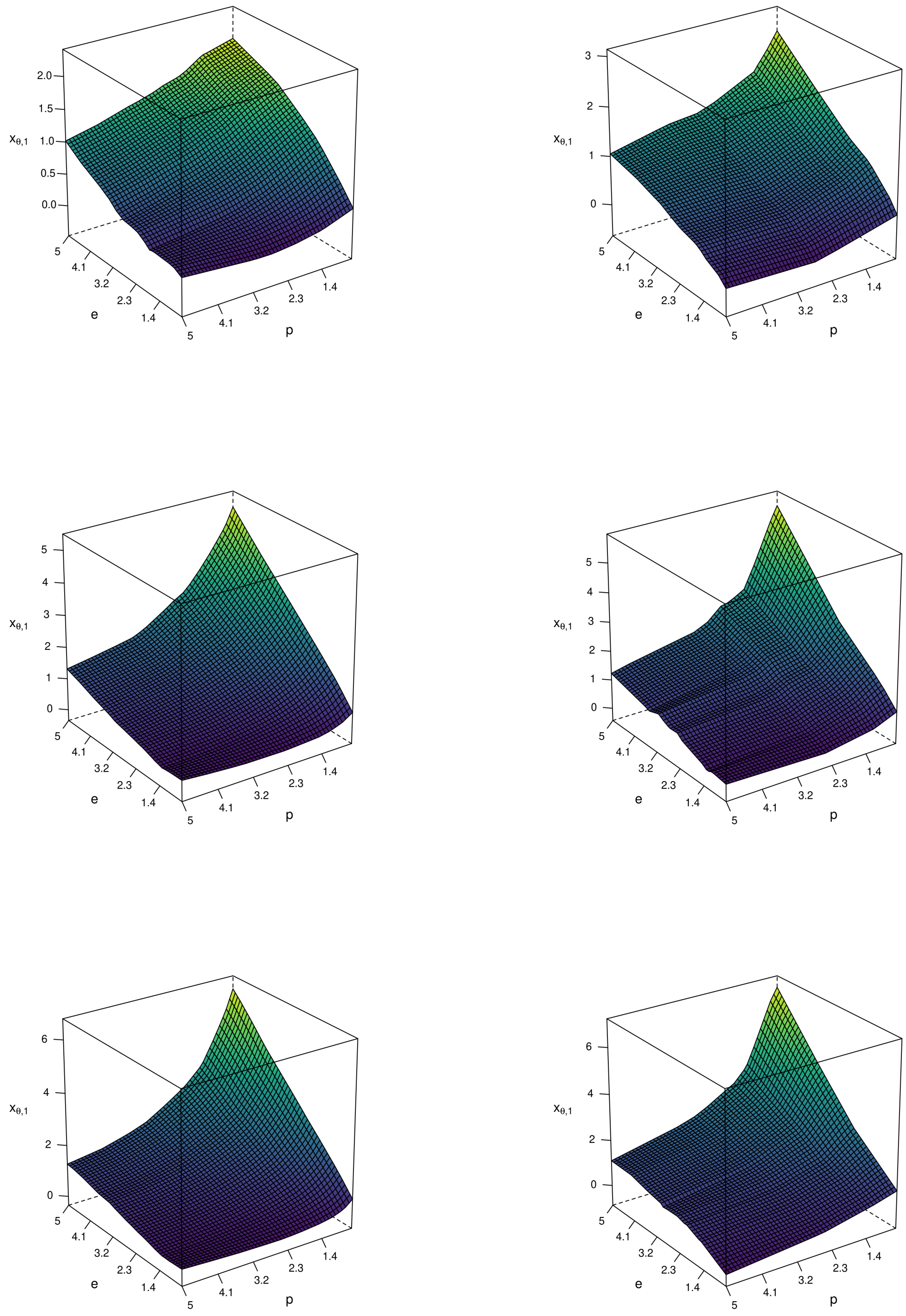

Figure 1.29. Estimates of conditional quantiles of food consumption given varous levels of τ. From top to bottom, I plot, in order, $\tau = 0.10, 0.75, 0.90$. In each row, on the left, I plot the estimate for the full NHCP sample, and on the right, I plot the estimate for the subset of the NHCP containing poorer households with more members and younger female heads.

Chapter 2

Revealed Stochastic Preference

The estimation procedure in Chapter 1 assumes that there exists a conditional quantile of consumption that coincides with an individual demand function. In demand analysis, it is common to assume that there is a one-to-one relationship between demand (in the absence of food stamps) and preferences, and that *every* conditional quantile of consumption is an individual demand function. This assumption is extremely strong because it assumes that heterogeneity in the population is finite-dimensional, and that Engel curves do not cross. I use a variant of this assumption in Chapter 1 because it is simple and tractable, and existing more flexible methods are lacking.

In this chapter, random fields are used to develop a non-parametric model of consumption with quasi-rational consumers and infinite-dimensional heterogeneity that is appropriate for scanner data. This model is used to recover the latent distribution of preferences in the population and perform counterfactual analysis. If variation in preferences is small, preferences can be recovered by approximating the relationship between demand and preferences using a first-order expansion. Else, preferences can be recovered numerically. This model is also used to analyze the behaviour of a representative consumer. In particular, I provide a test for the integrability of the expected demand field at a parametric rate, and recover the preferences associated with this field. The methods constructed in this chapter are illustrated in an application to the consumption of alcohol using the Nielsen Homescan Consumer Panel (NHCP).

2.1 Introduction

Scanner datasets have recently become easily accessible. These datasets contain detailed information on the price, quantity, and date of every purchase made by a large number of consumers. The name follows from the way that this information is collected: Either retailers or households record this information by "scanning" purchased goods.

The aim of this paper is to develop a model of consumption for performing demand analysis with scanner data, without imposing strong restrictions on preferences. This chapter is one of the first analyses on this topic (also see Burda et al., 2008, 2012, Guha and Ng, 2019, and Chernozhukov et al., 2020).

Standard approaches are not appropriate for scanner data. Parametric approaches impose unrealistic functional forms and error structures (see Brown and Walker, 1989, and Lewbel, 2001); existing non-parametric approaches require all consumers to face the same prices[1] (Blundell, Kristensen, and Matzkin, 2017), lack point identification (Dette et al., 2016; Hausman and Newey, 2016a), and/or require demand fields to be monotonic in low-dimensional heterogeneity (Blundell, Horowitz, and Parey, 2017), a strong restriction that requires individual demand fields to stack on top of each other (see Chapter 1).

This chapter develops a non-parametric approach without these limitations. In particular, infinite-dimensional preference heterogeneity is introduced by replacing the marginal rate of substitution field with a *random field* (see Gorman, 1953, 1961, for some of the first uses of fields in the demand literature, and Beckert and Blundell, 2008, and de Clippel and Rozen, 2020, for related uses of the marginal rate of substitution field as a baseline for modelling). Prices can vary significantly across consumers. Point identification follows from the panel structure of scanner data, the random field assumption, and an assumption on the way that preferences map to consumption. To my knowledge, this analysis is the first to replace the marginal rate of subsitution with a random field.

Consumers are assumed to be *quasi-rational*: Even if a realization of the marginal rate of substitution random field is not compatible with a well-behaved utility function, the consumer solves a system that is analogous to the first-order conditions in a standard utility maximization problem. Each realization of the marginal rate of substitution random field can be thought of as an implicit relative valuation that the consumer equates with relative prices. That being said, some random fields will almost surely produce field trajectories that can be interpreted as the marginal rate of substitution fields of well-behaved utility functions. In either case, demand in the population is also a random field.

This model can be used to recover the distributions of demand and preferences in the population, and perform individual-level counterfactual analysis. It can also be used to test the integrability of the demand field of a "representative consumer," and recover the preferences of this consumer.

Recoverability solves an ill-posed problem. When variation in preferences is small, preferences can be recovered by approximating the relationship between the marginal rate of substitution and demand random fields using a first-order expansion, and applying an analogue of the delta-method. This approximation yields the explicit form of conditional heteroscedasticity on the demand random field. When variation in preferences is large, preferences can be recovered numerically. The model is easy to simulate, avoiding several common computational issues.

In this framework, counterfactual analysis involves a state-space representation and a type of Kalman-filter algorithm (see Wikle and Cressie, 1999). Although, in many non-parametric settings, we can only recover wide bounds (see Varian, 1982), the fact that the preferences are drawn from a distribution lets us make more accurate predictions.

[1]Scanner datasets have a significant amount of price heterogeneity (see Section 2.6).

In general, there are two conditions to check for integrability: the symmetry of the Slutsky matrix (necessary for the expected demand field to satisfy the standard first-order conditions), and the negative semi-definiteness of the Slutsky matrix (necessary for demand to be consistent with utility maximization). In the current framework, symmetry is automatically satisfied. I test the integrability of the expected demand field by testing a condition that is slightly stronger than negative semi-definiteness (see Samuelson, 1948, 1950, Hurwicz and Uzawa, 1971, and Hosoya, 2013, 2016, for more on integrability).

The test is based on the theory of generalized functions. It is constructed by interpreting the integrability condition as a restriction of functionals. This interpretation lets us avoid estimating partial derivatives, in order to achieve a parametric rate (see Lewbel, 1995, and Haag et al., 2009, for two related tests of Slutsky symmetry).

The rest of the paper considers a two-good framework, and is organized as follows: In Section 2.2, I briefly review the standard consumer theory for a twice-continuously-differentiable, strictly increasing, and strongly quasi-concave utility function in the case of two goods. I focus on (i) the implicit equations that relate the marginal rate of substitution field, the demand field, and the indifference curves, and (ii) a positivity condition on the bordered Hessian of the utility function (concerning the curvature of the indifference curves) and a negativity condition on the "Slutsky coefficient" (concerning the demand function). In Section 2.3, I describe the *stochastic marginal rate of substitution* (SMRS) model, and explain how to deduce both the demand random field and the stochastic indifference curves from the marginal rate of substitution random field. In Section 2.4, I describe (i) the distributional assumptions on the observations, (ii) how to identify and estimate the demand random field, (iii) how to identify and estimate the marginal rate of substitution random field, and (iv) how to perform individual-level counterfactual analysis. In Section 2.5, I describe how to test the integrability of the demand field of a representative consumer, and recover the preferences of this consumer (when integrability is satisfied). Section 2.6 illustrates the approach with an application to alcohol consumption using scanner data from the Nielsen Homescan Consumer Panel (NHCP). Section 2.7 concludes. Proofs, summary statistics, and a number of technical details are placed in the appendix. Additional details (including a description of the underlying probability space) are also placed in the appendix.

2.2 Consumer Theory

In this section, I briefly review selected results in consumer theory (see Varian, 1992, Barten and Böhm, 1993, and Mas-Colell et al., 1995, for broad presentations). I discuss the indifference curves, the marginal rate of substitution (MRS), the demand field, and their relationships. I place an emphasis on the integrability condition that is needed to be able to recover the MRS from the demand field by inversion.

2.2.1 Preferences

Suppose that there are two distinct goods. Let $\bar{R} = \mathbb{R}^2_+$ denote the non-negative orthant with interior R. A consumer has preferences over bundles of goods $x \in \bar{R}$. Her preferences are summarized by a utility function $u : \bar{X} \to \mathbb{R}$. Let $G(v) = \{x \in \bar{X} : u(x) = v\}$ denote the equivalence class containing all bundles that attain a utility level of $v \in \mathbb{R}$.

Consider the following assumption on the utility function $u(\cdot)$ (as in Chapter 1):

Assumption 2.1.

(i) Utility $u(\cdot)$ is twice-continuously-differentiable on R.

(ii) Utility $u(\cdot)$ is strictly increasing with strictly positive partial derivatives on R.

(iii) Utility $u(\cdot)$ is strongly quasi-concave on R:

$$\xi' \frac{\partial^2 u(x)}{\partial x \partial x'} \xi < 0,$$

for every $\xi \in \mathbb{R}^2$ such that $\xi \neq 0$ and $\xi' \frac{\partial u(x)}{\partial x} = 0$, at each $x \in R$.

(iv) For each $v \in \mathbb{R}$ such that $v \neq u(0)$, $G(v)$ is contained in R.

Under Assumption 2.1(i), preferences are smooth (see Proposition 2.3.9 in Mas-Colell, 1985). Assumption 2.1(ii) implies that *more is strictly better*. Assumption 2.1(iii) implies that *averages are strictly better*.[2] Assumption 2.1(iv) implies that the boundary of $\bar{R}$ is undesirable. Assumptions 2.1(iii) and 2.1(iv) are common (see pages 415-416 in Katzner, 1968, and Sections 11-12 in Barten and Böhm, 1993).

Under Assumptions 2.1(i) to 2.1(ii), $G(v)$ is characterized by an indifference curve:

$$G(v) = \big\{x \in \bar{R} : x_2 = g(x_1, v)\big\}.$$

The collection of curves $g(\cdot)$ defines a field, indexed by (x_1, v). This field satisfies the following equation:

$$u(x_1, g(x_1, v)) = v, \tag{2.2.1}$$

for every $x_1 > 0$ and $v \neq u(0, 0)$. Therefore, the implicit function theorem implies that the indifference curve is twice-continuously-differentiable with respect to x_1 such that:[3]

$$\frac{\partial g(x_1, v)}{\partial x_1} = -\frac{u_1(x_1, g(x_1, v))}{u_2(x_1, g(x_1, v))} = -m(x_1, g(x_1, v)), \tag{2.2.2}$$

where $m(x) \equiv u_1(x)/u_2(x)$ denotes the *marginal rate of substitution* (MRS) at x, i.e. the rate at which

[2] Strong quasi-concavity is actually a slightly stronger condition that is characterized below.
[3] Throughout, $u_i(x)$ denotes $\partial u(x)/\partial x_i$, and $u_{ij}(x)$ denotes $\partial^2 u(x)/\partial x_i \partial x_j$.

the consumer is willing to exchange good 1 for good 2 given x. The marginal rate of substitution defines a second field, indexed by x.

Lemma 2.1. Under Assumptions 2.1(i), 2.1(ii) and 2.1(iv), the following properties are equivalent:

(i) The utility function $u(\cdot)$ is strongly quasi-concave on R.

(ii) The determinant of the bordered Hessian of $u(\cdot)$ is strictly positive:

$$\Delta_u(x) \equiv \det \begin{pmatrix} 0 & u_1(x) & u_2(x) \\ u_1(x) & u_{11}(x) & u_{12}(x) \\ u_2(x) & u_{21}(x) & u_{22}(x) \end{pmatrix} > 0, \tag{2.2.3}$$

at every $x \in R$.

(iii) The following function of the marginal rate of substitution is strictly positive:

$$\Delta_m(x) \equiv -\frac{\partial m}{\partial x_1}(x) + m(x)\frac{\partial m}{\partial x_2}(x) > 0, \tag{2.2.4}$$

at every $x \in R$.

(iv) The indifference curve is strictly convex with respect to x_1:

$$\frac{\partial^2 g(x_1, v)}{\partial x_1^2} > 0, \tag{2.2.5}$$

at every $x_1 > 0$ and $v \neq u(0,0)$.

Proof. See Appendix 2.A.1. □

Lemma 2.1 says that strong quasi-concavity is equivalent to a condition on the sign of the determinant of the bordered Hessian of $u(\cdot)$, which is equivalent to a diminishing marginal rate of substitution, and to a positivity condition on the second derivative of the indifference curve $g(\cdot, v)$. Strong quasi-concavity implies strict quasi-concavity, a common assumption in economics that requires the upper contour sets of the utility function to be strictly convex (Ginsberg, 1973), but the reverse implication does not hold because strict quasi-concavity permits the indifference curves to have zero curvature on a nowhere dense set (see Katzner, 1968).

While there is some difficulty in interpreting the positivity of $\Delta_m(\cdot)$, this restriction has a natural interpretation for homothetic preferences. If preferences are homothetic, then the marginal rate of substitution field $m(\cdot)$ has the form:

$$m(x) = m_0\left(\frac{x_2}{x_1}\right),$$

for some univariate function $m_0(\cdot)$, and every $x \in R$. The function in (2.2.4) equals:

$$\Delta_m(x) = -m_0'\left(\frac{x_2}{x_1}\right)\left[m_0\left(\frac{x_2}{x_1}\right)\frac{x_2}{x_1^2} + \frac{1}{x_1}\right],$$

for every $x \in R$. This quantity is strictly positive if, and only if, the function $m_0(\cdot)$ is strictly increasing. Intuitively, the rate at which the consumer is willing to exchange good 1 for good 2 given x must be strictly increasing in the slope x_2/x_1.

2.2.2 Demand Field

Let $\tilde{p}_j$ denote the price of good j, for each $j = 1, 2$. The consumer has income $\tilde{y} > 0$. She can afford a bundle $x \in \bar{R}$, if $\tilde{p}_1 x_1 + \tilde{p}_2 x_2 \leq \tilde{y}$. She chooses a bundle that solves:

$$\max_{x \in \bar{R}} u(x) \text{ subject to } \tilde{p}_1 x_1 + \tilde{p}_2 x_2 \leq \tilde{y}. \tag{2.2.6}$$

Since the solution to (2.2.6) is invariant to homothetic changes in prices and income, the second good can be made a numéraire, and prices and income can be measured relative to the price of the second good. The maximization problem in (2.2.6) becomes:

$$\max_{x \in \bar{R}} u(x) \text{ subject to } p x_1 + x_2 \leq y, \tag{2.2.7}$$

where $p = \tilde{p}_1/\tilde{p}_2$ is the relative price of good 1 and $y = \tilde{y}/\tilde{p}_2$ denotes income measured in units of good 2.

Let $x(z)$ denote the consumer's demand given the design $z = (y, p)$—that is, the set of bundles that solve (2.2.7). Under Assumption 2.1, a bundle $x \in \bar{R}$ is a member of $x(z)$ if, and only if, it solves the following system of first-order equations:

$$m(x) = p \text{ and } p x_1 + x_2 = y. \tag{2.2.8}$$

From the first equality, it follows that the slope of the indifference curve at the optimum equals the slope of the budget line. The second equality is Walras' law (Walras, 1874), which says that the optimum is on the boundary of the budget set, an implication of the monotonicity of the utility function $u(\cdot)$. Equivalently, one can solve:

$$m\left(x_1, y - p x_1\right) - p = 0, \tag{2.2.9}$$

for x_1, and then use the second equality in (2.2.8) to solve for x_2. This implicit equation characterizes demand. In particular, it explains how to deduce $x_1(z)$ from $m(\cdot)$. The implicit function theorem implies that demand is single-valued and continuously-differentiable on R (see Appendix 2.A.2). Thus, demand $x(\cdot)$ defines another field, indexed by z.

Strong quasi-concavity has the following implication:

Lemma 2.2. Under Assumption 2.1, the *Slutsky coefficient* (or *compensated price derivative*) is negative:

$$\Delta_x(z) \equiv \frac{\partial x_1}{\partial p}(z) + x_1(z)\frac{\partial x_1}{\partial y}(z) < 0, \tag{2.2.10}$$

for every $z \in R$.

Proof. See the proof of Theorem 13.1(iv) in Barten and Böhm (1993). □

Lemma 2.2 provides a testable implication of strong quasi-concavity. The Slutsky coefficient $\Delta_x(z)$ is proportional to the North-West entry of the 2-by-2 Slutsky matrix (see the summary of Slutsky, 1915, in Allen, 1936). When there are two goods:

(i) the Slutsky matrix is automatically symmetric;

(ii) the Slutsky matrix is negative semi-definite if, and only if, $\Delta_x(z) \leq 0$;

(iii) exactly one eigenvalue of the Slutsky matrix is strictly negative if, and only if, $\Delta_x(z) < 0$.

A proof of these relationships is placed in Appendix 2.A.3. The fact that the inequality in (2.2.10) is strict is essential for the proof of Proposition 2.1 below, and ultimately, for the identification of preferences, described in Section 2.2.3. If strong quasi-concavity were replaced with strict quasi-concavity, then the Slutsky matrix would be negative semi-definite, ensuring that $\Delta_x(z)$ were non-positive, but not necessarily strictly negative.

Lemma 2.2 itself has the following implication:

Proposition 2.1. Under Assumption 2.1, the following properties hold:

(i) Demand $x(\cdot)$ is invertible on R.

(ii) Inverse demand, denoted $z(\cdot) = \left[y(\cdot), p(\cdot)\right]'$, is continuously-differentiable on R.

(iii) The following relationship holds on R:

$$\Delta_x(z) = -\Delta_m(x(z))^{-1}, \tag{2.2.11}$$

for every $z \in R$.

Proof. See Appendix 2.A.4. □

Proposition 2.1(iii) implies that, under Assumptions 2.1(i), 2.1(ii) and 2.1(iv), *strong quasi-concavity is equivalent to the negativity of* $\Delta_x(z)$, which is equivalent to "simple concavity" used for the integrability in Samuelson (1948).[4]

[4]This result follows from (2.2.8). See the footnote on page 243 of Samuelson (1948).

Inverse demand $z(\cdot)$ defines a fourth field, indexed by x. Because the demand field $x(\cdot)$ satisfies the first-order conditions in (2.2.8), the inverse demand field $z(\cdot)$ satisfies:

$$p(x) = m(x) \text{ and } y(x) = m(x)x_1 + x_2, \tag{2.2.12}$$

for every $x \in R$. Therefore, there is a one-to-one relationship between inverse demand $z(\cdot)$ and the marginal rate of substitution $m(\cdot)$.

2.2.3 Identification

In Section 2.2.2, we showed how to deduce the demand field $x(\cdot)$ from the marginal rate of substitution field $m(\cdot)$. Let us now show how to deduce the marginal rate of substitution field $m(\cdot)$ from the demand field $x(\cdot)$, and the field of indifference curves $g(\cdot)$ from the marginal rate of substitution field $m(\cdot)$—that is, the procedure described below "identifies" preferences from the observation of the demand field $x(\cdot)$.[5]

First, consider the following observability assumption:

Assumption A.

(i) Demand $x(\cdot)$ is observed on $\mathcal{Z} \subseteq R$.

(ii) The closure $\mathcal{X}$ of the range $x(\mathcal{Z})$ admits an open subset $\mathcal{X}_0$ that is dense in $\mathcal{X}$.

Under Assumption A, the demand field $x(\cdot)$ is observed on a subset $\mathcal{Z}$ of the interior R, and there exists an open subset $\mathcal{X}_0$ of the range $x(\mathcal{Z})$ whose closure $\mathcal{X}$ coincides with the closure of this range. Assumption A is less stringent than the usual assumption that the demand field $x(\cdot)$ is observed on R.

Now, let:

$$W_v(\mathcal{X}) = \{x_1 \geq 0 : (x_1, x_2) \in \mathcal{X} \cap G(v), \text{ for some } x_2 \geq 0\}, \tag{2.2.13}$$

denote the projection of the intersection of the closure of the range $x(\mathcal{Z})$ and the graph of the indifference curve $g(\cdot, v)$ associated with a utility level $v \in \mathbb{R}$ onto the x_1-axis.

We obtain the following result:

Theorem 2.1. Under Assumptions 2.1 and A:

(i) The marginal rate of substitution $m(\cdot)$ is identified over $\mathcal{X}$.

(ii) If $W_v(\mathcal{X})$ is connected, the curve $g(\cdot, v)$ is identified over the interior of $W_v(\mathcal{X})$.

Proof. See Appendix 2.A.5. □

Theorem 2.1(i) implies that the marginal rate of substitution field $m(\cdot)$ is identified over the closure of the range of demand. This result does not rely on the restrictions on $\mathcal{X}$ in Assumption A(ii). By

[5] Identification is related to integrability and recoverability in consumer theory (see Mas-Colell, 1977). The discussion in this section provides a brief, more formal discussion of the results in Section 1.3.4 in Chapter 1.

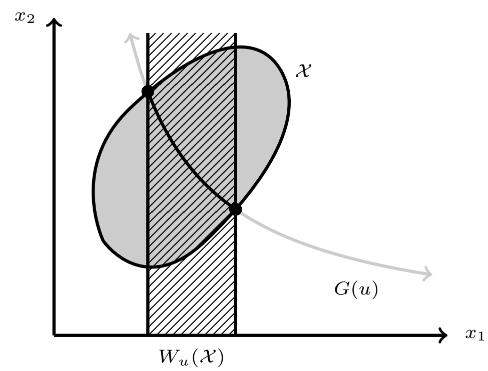

Figure 2.1. This figure illustrates the sets from Theorem 2.1. The shaded set is the range of demand $\mathcal{X}$. I draw an indifference curve whose intersection with $\mathcal{X}$ is connected, so that $W_u(\mathcal{X})$ is connected.

Theorem 2.1(ii), the indifference curve $g(\cdot, v)$ is identified wherever it intersects $\mathcal{X}$, as long as it does not leave and then re-enter $\mathcal{X}$ (see Figure 2.1). Theorem 2.1(ii) follows immediately from Theorem 2.1(i) and the Picard-Lindelöf theorem applied to the differential equation in (2.2.2) subject to the initial condition $g(x_1^*, v) = x_2^*$, for any observable bundle $x^* \in G(v)$. The restrictions on $\mathcal{X}$ in Assumption A(ii) ensure that the solution to this differential equation can be extended to the boundary of $\mathcal{X}$. This result is related to the integrability theorem (see, for example, pages 243-245 in Samuelson, 1948, Theorem 2 in Hurwicz and Uzawa, 1971, Theorem 2 in Hosoya, 2013, and Section 2.4 in Hosoya, 2016), and obtained by assuming that the demand field $x(\cdot)$ is generated by a well-behaved utility function $u(\cdot)$. That being said, I only use this assumption to characterize the subset on which we can recover the indifference curve. Indeed, the strict negativity of the Slutsky coefficient $\Delta_x(\cdot)$ is necessary and sufficient for there to exist a unique recoverable utility function $u(\cdot)$.

Let us now consider the steps for recovering the consumer's preferences:

Step 1: Invert the demand field $x(\cdot)$ to recover the inverse demand field $z(\cdot)$ over $\mathcal{X}$. By the first equality in (2.2.12), the second component $p(\cdot)$ of this inverse is equal to the marginal rate of substitution field $m(\cdot)$.

Step 2: Fix an observable bundle $x^* \in \mathcal{X}$, and let $v \in \mathbb{R}$ denote the utility level atta- ined by x^*. Solve the ordinary differential equation in (2.2.2):

$$\frac{\partial g(x_1, v)}{\partial x_1} = -m(x_1, g(x_1, v)), \tag{2.2.14}$$

with respect to x_1 over the interior of $W_v(\mathcal{X})$ subject to $g(x_1^*, v) = x_2^*$. The solution coincides with the indifference curve that passes through bundle x^*.

Example 2.1. For illustration, let us consider a Stone-Geary utility function with equal weights, defined

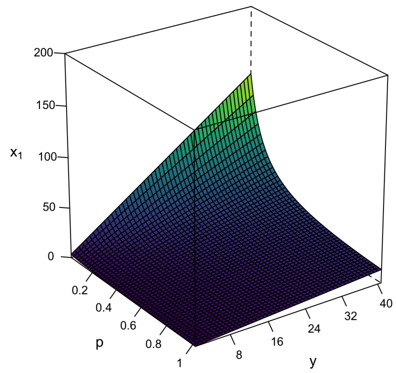

Figure 2.2. An illustration of $x_1(\cdot)$ in Example 2.3. I use Stone-Geary $\mu(\cdot)$, $k_1 = 10$, $k_2 = 2$, and $\sigma_1 = \sigma_2 = 1/2$. See Appendix 2.D for details.

by: $u(x) = x_1^{\frac{1}{2}} x_2^{\frac{1}{2}}$, for each $x \in \bar{R}$. Under this specification, we obtain:

$$m(x) = \frac{x_2}{x_1}. \tag{2.2.15}$$

By solving the first-order conditions in (2.2.8) and inverting:

$$x(z) = \left(\frac{y}{2p}, \frac{y}{2}\right)' \quad \text{and} \quad z(x) = \left(2x_2, \frac{x_2}{x_1}\right)', \tag{2.2.16}$$

verifying the equality of $m(\cdot)$ and $p(\cdot) \equiv z_2(\cdot)$. Figure 2.2 illustrates the first component of this demand field. For simplicity, let us assume that this demand field is observed on $\mathcal{Z} = R$. Under this assumption, the range $\mathcal{X} = R$ is open and convex, and $W_v(\mathcal{X})$ is the set of all positive numbers, for each $v > 0$. The differential equation in (2.2.2) becomes a linear differential equation such that:

$$\frac{\partial g(x_1, v)}{\partial x_1} = -\frac{g(x_1, v)}{x_1}. \tag{2.2.17}$$

The solution to this differential equation has the following form:

$$g_0(x_1, v) = \frac{\delta_v}{x_1}, \tag{2.2.18}$$

where the integrating constant δ_v depends on the utility level $v \in \mathbb{R}$. For a fixed utility level $v \in \mathbb{R}$, and a given initial condition $g(x_1^*, v) = x_2^*$, this solution is:

$$g_0(x_1, v) = \frac{x_1^* x_2^*}{x_1}. \tag{2.2.19}$$

This solution is the indifference curve that passes through x^* because $g(x_1, v) = \frac{u^2}{x_1}$ and $v = u(x^*)$ implies $v^2 = x_1^* x_2^*$. △

2.3 Stochastic Preferences and Demand

In this section, preferences are made stochastic by replacing the deterministic marginal rate of substitution field $m(\cdot)$ with a *random field* $M(\cdot)$. This approach differs from the practice of imposing a stochastic assumption onto the utility function (see Dette et al., 2016, Kitamura and Stoye, 2018, and Deb et al., 2018). Instead, I consider the marginal rate of substitution field to be the baseline for modelling (see Bansal and Yaron, 2004, for its use in an analysis of intertemporal consumption and asset pricing, Beckert and Blundell, 2008, for its use in a setting like ours, and de Clippel and Rozen, 2020, for its use in some recent literature on revealed preference). The marginal rate of substitution is used because it is a *cardinal* characterization of preferences. Consequently, we can expect to identify the distribution of preferences, rather than a class of distributions that is unique up to an increasing transformation. If we were to, instead, make the utility function stochastic, then we would also risk imposing assumptions with misleading interpretations. For instance, the independence of three utility levels $u(x)$, $u(\tilde{x})$, and $u(\check{x})$ has no economic significance because the utility levels 0, $u(\tilde{x}) - u(x)$, and $u(\check{x}) - u(x)$ lead to the same preference ordering, but violate independence. In this respect, the terminology "random utility" (introduced by Thurstone, 1927, Domencich and McFadden, 1975, and McFadden, 1981) is misleading. I refer to the model in this section a "stochastic marginal rate of substitution" (SMRS) model. In this framework, for each realization of the marginal rate of substitution random field, an analogue of the first-order conditions in (2.2.9) is used to deduce demand. Hence, demand is also a random field $X(\cdot)$.

2.3.1 Marginal Rate of Substitution Random Field

Fix a probability space $(\Omega, \mathcal{A}, P)$.[6] The deterministic field $m(\cdot)$ is replaced with a stochastic function $M(\cdot)$ such that $M(x) = M(x; \omega)$, where $\omega \in \Omega$. In particular, I consider a Gaussian field for $\log M(\cdot)$ to treat the positivity of the marginal rate of substitution $M(\cdot)$.[7] This chapter appears to be the first to model stochastic preferences by replacing the marginal rate of substitution with a random field. Random fields have been used in economics to model interest rates (Kennedy, 1994; Goldstein, 2000), Arrow-Debreu prices (Clement et al., 2000), travel flows (Bolduc et al., 1992), peer effects (Lin, 2005, 2010), future ambiguity (Izhakian, 2020), networks (Leung, 2015; Boucher and Mourifié, 2017), and labour supply (Crawford, 2019). Random fields are also pervasive in other fields with applications in weather forecasting, disease and information diffusion, and facial recognition (Cressie, 1993; Vanmarcke, 1983).

Consider the following assumption:

[6] See Section 2.3.2 and Appendix 2.E for potential probability spaces.

[7] A stochastic field is called a spatial process when the index is discrete and a random field when the index is continuous.

Assumption 2.2.

(i) The random field $\log M(\cdot)$ has the form:

$$\log M(x) = \mu(x) + \sigma_1 U_1(x_1) + \sigma_2 U_2(x_2), \tag{2.3.1}$$

for every $x \in R$.

(ii) The processes, $U_1(\cdot)$ and $U_2(\cdot)$, are independent, univariate Gaussian processes.

(iii) The processes, $U_1(\cdot)$ and $U_2(\cdot)$, are zero-mean with unit-variance.

(iv) The processes, $U_1(\cdot)$ and $U_2(\cdot)$, are *stationary*.[8]

Assumption 2.2 implies that the random field $\log M(\cdot)$ is Gaussian with mean $\mu(x)$ and covariance operator:

$$C(x, \tilde{x}) = \sigma_1^2 C_1(x_1, \tilde{x}_1) + \sigma_2^2 C_2(x_2, \tilde{x}_2), \tag{2.3.2}$$

for every $x, \tilde{x} \in R$, where $C_j(\cdot)$ denotes the covariance operator for $U_j(\cdot)$ such that:

$$C_j(x_j, \tilde{x}_j) = \text{cov}(U_j(x_j), U_j(\tilde{x}_j)), \tag{2.3.3}$$

for $j = 1, 2$. The expression for the covariance operator $C(\cdot)$ follows from the independence of $U_1(\cdot)$ and $U_2(\cdot)$. Under Assumption 2.2, $\log M(\cdot) - \mu(\cdot)$ is strictly stationary. While Assumption 2.2 restricts the distribution of the errors to be Gaussian and $C(\cdot)$ to be separable, it does not restrict the mean, and, in general, $\log M(\cdot)$ does not satisfy mean stationarity. Assumption A2(iii) is simply a normalization that is made without loss of generality, implying that (i) $C_j(x, x) = 1$, for $j = 1, 2$, and (ii) $C(x, x) = \sigma_1^2 + \sigma_2^2$.

Assumption 2.2 can be used to deduce the distributional properties of $M(\cdot)$:

Proposition 2.2. Under Assumption 2.2:

(i) The random field $M(\cdot)$ is log-normal.

(ii) The expected marginal rate of substitution $\tilde{m}(\cdot)$ is:

$$\tilde{m}(x) \equiv \mathbb{E}[M(x)] = \exp\left\{\mu(x) + \frac{1}{2}\left[\sigma_1^2 + \sigma_2^2\right]\right\}. \tag{2.3.4}$$

(iii) The random field $M(\cdot)$ has a covariance operator $C_M(\cdot)$ of the form:

$$C_M(x, \tilde{x}) = \exp\left\{\mu(x) + \mu(\tilde{x}) + \sigma_1^2 + \sigma_2^2\right\}\left[\exp\{C(x, \tilde{x})\} - 1\right]. \tag{2.3.5}$$

Proof. See Appendix 2.A.6 ◻

[8] $U_j(\cdot)$ is stationary if the value of its covariance operator $C_j(\cdot)$ at $(x_j, \tilde{x}_j)$ only depends on $x_j - \tilde{x}_j$.

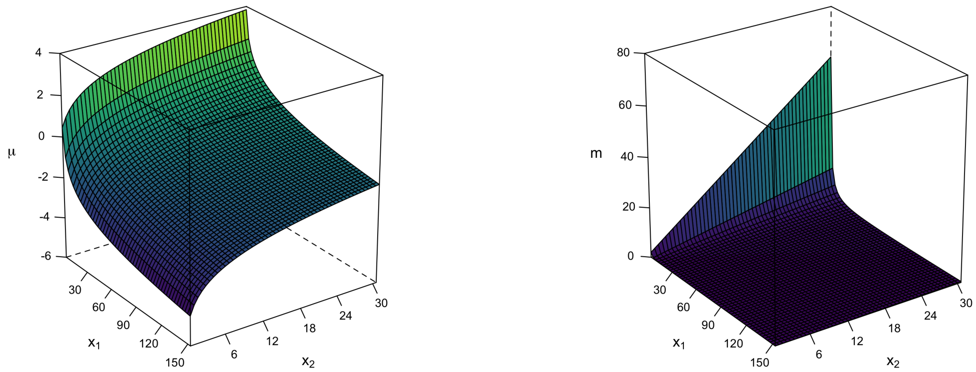

Figure 2.3. The surface on the left is $\mu(\cdot)$ in Example 2.2. The surface on the right is $\tilde{m}(\cdot)$ in Example 2.2. I use $\lambda_m = \log(2)$ and $\sigma^2 = 1/2$.

Because a log-normal field is characterized by the mean and covariance operator of the underlying Gaussian field, the marginal rate of substitution random field $M(\cdot)$ is characterized by the functional parameters, $\mu(\cdot)$ and $C(\cdot)$, and the scalar parameters (σ_1, σ_2). Equivalently, it is characterized by the expected marginal rate of substitution $\tilde{m}(\cdot)$, the operator $C(\cdot)$, and the scalar parameters (σ_1, σ_2), since there is a one-to-one relationship between $\mu(\cdot)$ and $\tilde{m}(\cdot)$ given $C(\cdot)$ and (σ_1, σ_2).

Consider three examples:

Example 2.2. Consider a simple parametric model with scalar heterogeneity to ground the reader's perception of this framework in something familiar. Specifically, consider the general form of the Stone-Geary specification in Example 2.1, defined by: $u(x) = x_1^{\alpha} x_2^{1-\alpha}$, for some $\alpha \in (0,1)$ and each $x \in \bar{R}$. Under this specification, we obtain:

$$m(x) = \left(\frac{\alpha}{1-\alpha}\right)\frac{x_2}{x_1}, \tag{2.3.6}$$

for every $x \in R$. Scalar heterogeneity is introduced by assuming that $\lambda \equiv \frac{\alpha}{1-\alpha}$ is log-normal. Under this assumption:

$$\log M(x) = \log(x_2) - \log(x_1) + \lambda_m + \sigma^2 U, \tag{2.3.7}$$

in which U is standard normal and λ_m denotes the mean of $\log \lambda$. This model satisfies Assumption 2.2 with (i) $\mu(x) = \log(x_2) - \log(x_1) + \lambda_m$, (ii) $\sigma_1 = \sigma$, (iii) $\sigma_2 = 0$, and (iv) $U_1(x_1) = U$. Under this specification, Proposition 2.2 implies:

$$\tilde{m}(x) = \frac{x_2}{x_1} \cdot e^{\lambda_m + \frac{\sigma^2}{2}}. \tag{2.3.8}$$

Therefore, the expected marginal rate of substitution $\tilde{m}(\cdot)$ is the Stone-Geary marginal rate of substi-

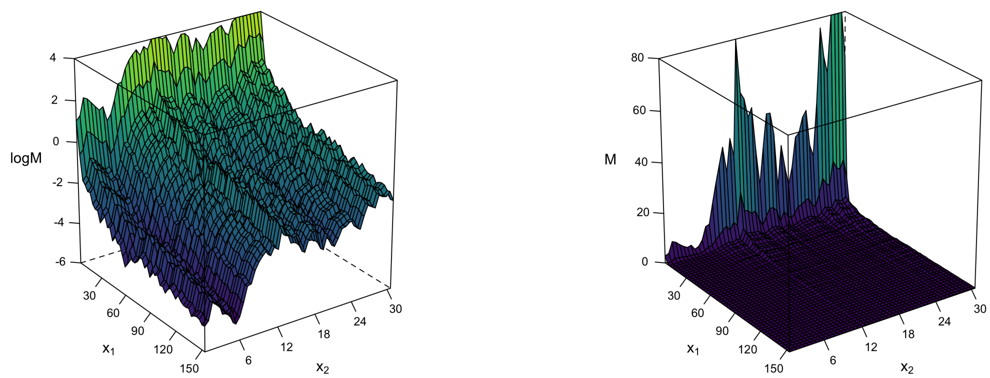

Figure 2.4. The surface on the left is $\log M(\cdot)$ under Ornstein-Uhlenbeck errors. The surface on the right is $M(\cdot)$ under Ornstein-Uhlenbeck errors. I use $\mu(x) = \log(x_2) - \log(x_1)$, $k_1 = 10$, $k_2 = 2$, and $\sigma_1 = \sigma_2 = 1/2$. See Appendix 2.D for details.

tution given the parameter:

$$\alpha = \left(1 + e^{-\lambda_m - \frac{\sigma^2}{2}}\right)^{-1}. \tag{2.3.9}$$

△

Example 2.3. Suppose that $U_j(\cdot)$ is an independent Ornstein-Uhlenbeck diffusion process with drift $k_j > 0$ and volatility $\eta_j > 0$, for each $j = 1, 2$. The Ornstein-Uhlenbeck process $U_j(\cdot)$ is the stationary solution to the following stochastic differential equation (SDE):

$$dU_j(x_j) = -k_j U_j(x_j) dx_j + \eta_j dW_j(x_j), \tag{2.3.10}$$

where $W_j(\cdot)$ is a Brownian motion. This solution is a Gaussian process with zero-mean and covariance operator:

$$C_j(x_j, \tilde{x}_j) = \frac{\eta_j^2}{2k_j} \exp\{-k_j|x_j - \tilde{x}_j|\}. \tag{2.3.11}$$

To satisfy Assumption 2.2(iii), this process must have unit-variance. This property holds if, and only if, $\eta_j = \sqrt{2k_j}$. Under this restriction, we get a Gaussian process with zero-mean and covariance operator:

$$C_j(x_j, \tilde{x}_j) = \exp\{-k_j|x_j - \tilde{x}_j|\}. \tag{2.3.12}$$

If both $U_1(\cdot)$ and $U_2(\cdot)$ are defined in this way, then:

$$C(x, \tilde{x}) = \sigma_1^2 \exp\{-k_1|x_1 - \tilde{x}_1|\} + \sigma_2^2 \exp\{-k_2|x_2 - \tilde{x}_2|\}. \tag{2.3.13}$$

The zero-mean Ornstein-Uhlenbeck process (see Ornstein and Uhlenbeck, 1930) is the only strictly stationary, Gaussian, Markov process in one-dimension with zero-mean and continuous trajectories. This process is tractable because (i) it is easy to simulate, (ii) it is the analogue of a Gaussian AR(1) process in continuous space (to be precise, a space-discretized Ornstein-Uhlenbeck process is a Gaussian AR(1) process, for any discretization step), and (iii) it is a diffusion process. Figure 2.4 displays a realization of the marginal rate of substitution random field $M(\cdot)$ given $\mu(x) = \log(x_2) - \log(x_1)$ and Ornstein-Uhlenbeck errors with parameters $k_1 = 10$, $k_2 = 2$, and $\sigma_1 = \sigma_2 = 1/2$. $\triangle$

Example 2.4. While the Ornstein-Uhlenbeck process is tractable *because* it is the stationary solution of an SDE, Assumption 2.2 does not require $U_j(\cdot)$ to be a diffusion process. The Gaussian distribution of $U_j(\cdot)$ can be directly defined by its covariance operator. To illustrate, suppose that $U_j(\cdot)$ is an independent zero-mean Gaussian process with a squared-exponential covariance operator, parameterized by some $k_j > 0$:

$$C_j(x_j, \tilde{x}_j) = \exp\{-k_j|x_j - \tilde{x}_j|^2\}, \tag{2.3.14}$$

for each $j = 1, 2$. It is rather easy to show that this operator is positive semi-definite, making it a valid covariance operator. Under this specification:

$$C(x, \tilde{x}) = \sigma_1^2 \exp\big\{ - k_1|x_1 - \tilde{x}_1|^2\big\} + \sigma_2^2 \exp\big\{ - k_2|x_2 - \tilde{x}_2|^2\big\}. \tag{2.3.15}$$

The squared exponential covariance operator is the most used covariance operator in the machine learning literature (Rasmussen and Williams, 2006). Like the Ornstein-Uhlenbeck process in Example 2.3, this process is tractable because (i) it is easy to simulate, and (ii) it is the analogue of a Gaussian AR(∞) process in continuous space. Figure 2.5 displays a realization of the marginal rate of substitution random field $M(\cdot)$ given $\mu(x) = \log(x_2) - \log(x_1)$ and squared-exponential errors with parameters $k_1 = 10$, $k_2 = 2$, and $\sigma_1 = \sigma_2 = 1/2$. $\triangle$

Remark 2.1. The choice of the random field for the error term is due to the difficulty in defining processes with continuous indices. Loosely speaking, it is not appropriate to directly extend the notion of white noise and say that errors $\varepsilon(x)$, x varying, are independent and identically distributed with, say, a Gaussian distribution. Indeed, such an assumption would create a jump in the field trajectory, at every $x \in R$. It would be without practical meaning, impeding the development of an appropriate probabilistic theory, and making it impossible to use local analysis (see Section 2.3.2, as well as Appendix 2.E).

Remark 2.2. The uncertainty is not represented by a single random variable, but by a random field. It depends on the bundle $x \in R$, allowing for conditional heteroscedasticity. The uncertainty on demand will also be represented by a random field, rather than a scalar or a finite-dimensional shock (see Beckert and Blundell, 2008, Blundell, Horowitz, and Parey, 2017, and Allen and Rehbeck, 2019, for examples of such strong restrictions).

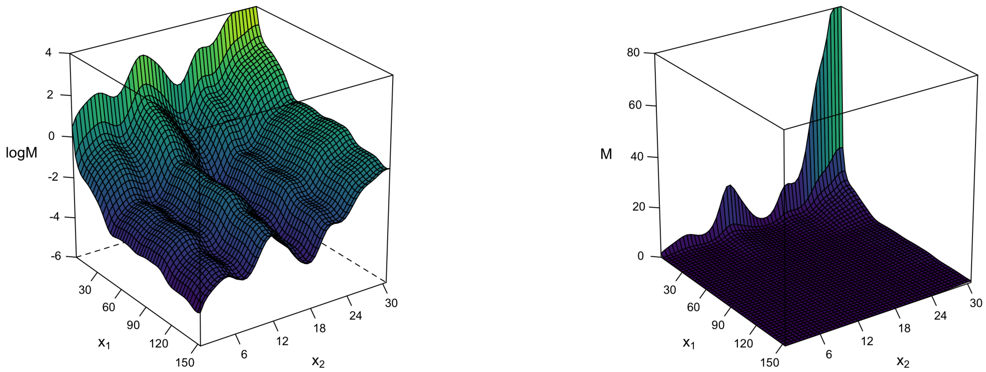

Figure 2.5. The surface on the left is $\log M(\cdot)$ under squared-exponential errors. The surface on the right is $M(\cdot)$ under squared-exponential errors. I use $\mu(x) = \log(x_2) - \log(x_1)$, $k_1 = 10$, $k_2 = 2$, and $\sigma_1 = \sigma_2 = 1/2$. See Appendix 2.D for details.

Remark 2.3. The additively decomposable form of the error term in Assumption A2(i) is for simplicity. This decomposable form assumes that a shock to the marginal rate of substitution at x_1 will persist across all values of x_2 given x_1, and a shock to the marginal rate of substitution at x_2 will persist across all values of x_1 given x_2, a form of behavioural separability in taste uncertainty.

Remark 2.4. The parameter σ_j measures the variability of the shocks in the direction of good j, describing the degree of uncertainty in that direction. As σ_1 and σ_2 approach zero, we approach the deterministic model. If all choices are made by one consumer, and the preferences associated with the exponential transform of $\mu(\cdot)$ are "rational," then we can interpret $\sqrt{\sigma_1^2 + \sigma_2^2}$, as a measure of the size of the consumer's deviation from economic rationality. While not entirely analogous, this remark is related to a subset of the economic literature concerned with such measures (Afriat, 1967; Houtman and Maks, 1985; Varian, 1990; Echenique et al., 2011; Dean and Martin, 2016).

Remark 2.5. In a recent set of literature on revealed preference (see Echenique et al., 2011, Deb et al., 2018, and Allen and Rehbeck, 2019), it is common to assume quasi-linear preferences—in particular, utility functions $u(\cdot)$ of the form: $u(x) = \varphi(x_1) + x_2$, where $\varphi : \mathbb{R}_+ \to \mathbb{R}$ is a locally non-satiated function. If $\varphi(\cdot)$ is differentiable, then $m(x) = d\varphi(x_1)/dx_1$ does not depend on x_2, and $x_1(z)$ does not depend on y. In other words, quasi-linear preferences assume that there is no income effect on the demand for good 1. Even if such a restriction is placed on the mean $\tilde{m}(\cdot)$ of the marginal rate of substitution random field $M(\cdot)$, an income effect would appear in this random field because of the uncertainty, as long as $\sigma_1 \neq 0$.

2.3.2 Skorokhod Space

I have not yet discussed the conditions that are needed to define a random field. In particular, I have not defined a probability space. From now on, I restrict this space to be the Skorokhod space, commonly denoted by $D[0,c)^2$, where $c > 0$ can be finite or infinite. The Skorokhod space is standard for random fields, consisting of all surfaces that are *cadlag* (a French terminology that means right-continuous with left limits). This space can be turned into a complete separable metric space (also known as a Polish space) by equipping it with an appropriate metric (see McFadden, 2005, for the definition of a Polish space in consumer theory, and Gikhman and Skorokhod, 1966, Billingsley, 1999, and Appendix 2.E for examples of metrics). Therefore, the standard measurable space is $(\Omega, \mathcal{A}) = \big(D[0,c)^2, \mathscr{D}\big)$, where $\mathscr{D}$ denotes the Borel sets associated with the Skorokhod Polish space $D[0,c)^2$. This space is appropriate for one-dimensional stochastic surfaces such as $M(\cdot)$. In what follows, we will also have to consider two-dimensional stochastic surfaces such as the demand random field. For such surfaces, the standard measurable space is the product space $\big(D[0,c)^2, \mathscr{D}\big)^{\otimes 2}$. I introduce these spaces to apply extensions of standard results to random fields such as differentiability, weak convergence (convergence in distribution), the Continuous Mapping Theorem, and the Central Limit Theorem (see Appendix 2.E).

Let us now discuss consequences of the definition of the probability space: Without additional structure, there is no reason to believe that any particular realization of the marginal rate of substitution random field $M(\cdot)$ will be consistent with consumer theory (see also the discussion in Dette et al., 2016, especially footnote 5). For example, the trajectories of the Ornstein-Uhlenbeck process in Example 2.3 are non-differentiable *almost everywhere*, implying that the realizations of $M(\cdot)$ cannot necessarily be structurally interpreted as marginal rate of substitution fields. Next, consider the Gaussian process associated with the squared-exponential covariance operator in Example 2.4. The trajectories of this process have derivatives of all orders, ensuring that the ordinary differential equation in (2.2.2) can always be solved to obtain a unique family of indifference curves associated with *some* strictly increasing utility function, but the stochastic version of $\Delta_m(\cdot)$ will almost surely violate its positivity condition.

These inconsistencies with consumer theory will be resolved in two different ways: First, consumers will be assumed to be *quasi-rational* (see Assumption 2.3 in Section 2.3.3 below). The realizations of the marginal rate of substitution random field $M(\cdot)$ will be seen as implicit relative valuations that the consumer equates with relative prices. When $M(\cdot)$ is consistent with consumer theory, she solves the first-order conditions in (2.2.8), and when it is not, she solves an analogue of these conditions, without them being literally interpreted as first-order conditions. Second, even if the realizations of $M(\cdot)$ are not consistent with consumer theory, there can exist a rational *representative consumer*. In other words, expected demand might satisfy integrability, allowing us to construct and recover a coherent notion of aggregate preferences to carry out welfare analysis (see Section 2.5.1).

2.3.3 Demand Random Field

The marginal rate of substitution random field $M(\cdot)$ is not directly observed. Instead, we observe choices made by consumers under various designs $z = (y, p)$. Therefore, it is important to be precise about the relationship between the marginal rate of substitution random field $M(\cdot)$ and the demand random field, denoted $X(\cdot)$.

As mentioned above, even if a realization of the marginal rate of substitution random field $M(\cdot)$ is not compatible with consumer theory, it is assumed that the consumer solves a system that is analogous to the first-order conditions in (2.2.8). Without more structure, there is, a priori, no reason for there to exist a unique bundle that equates the marginal rate of substitution $M(\cdot)$ with the price p. Below, it is assumed that, regardless of the interpretation, the choice of the solution is done in a measurable way.

Assumption 2.3. The random field $X(\cdot)$ is a measurable solution to:

$$M(X_1(z), X_2(z)) = p \text{ and } pX_1(z) + X_2(z) = y. \tag{2.3.16}$$

As mentioned in Section 2.2.2, instead of solving the system in (2.3.16), one can solve:

$$M(X_1(z), y - pX_1(z)) - p = 0, \tag{2.3.17}$$

for $X_1(z)$, then use the second equality in (2.3.16) to solve for $X_2(z)$. Under Assumption 2.3, there is a measurable function $A(\cdot)$ that transforms the inverse demand random field, say $Z(\cdot)$, into the demand random field $X(\cdot)$. The properties of $X(\cdot)$ cannot be deduced in closed-form from the properties of $M(\cdot)$, since $A(\cdot)$ is in general non-linear.

2.3.4 Small-Sigma Approximation

If σ_1 and σ_2 are small, the implicit equation in (2.3.17) can be approximated by a first-order expansion. This method requires an appropriate definition of a differential (see Appendix 2.E.6). Under standard regularity conditions, an analogue of the delta-method (or Slutsky method) can be applied to the random fields to approximate the distributions of the marginal rate of substitution and demand random fields.

Consider the following assumption:

Assumption S. The parameters σ_1 and σ_2 are *sufficiently small*.

With slight abuse of notation, let $x(z) = \mathbb{E}[X(z)]$ denote the expectation of the demand random field given $z \in R$, and let $m(\cdot)$ denote the second component of its generalized inverse $\operatorname{inv} x(\cdot)$. If $x(\cdot)$ is a *bona fide* demand field, then $m(\cdot)$ is the marginal rate of substitution for this field, and the analogue of the differential equation in (2.2.2) can be solved to recover a unique field of indifference curves $g_0(\cdot)$. If a realization of $M(\cdot)$ is continuously-differentiable, then it also yields a unique preference field $G_0(\cdot)$, defined by the following ordinary differential equation:

$$\frac{\partial G_0(x_1, v)}{\partial x_1} = -M(x_1, G_0(x_1, v)). \tag{2.3.18}$$

For exposition, it is assumed, without loss of generality, that $g_0(\cdot, v)$ and $G_0(\cdot, v)$ pass through a bundle, denoted (x_{10}, x_{20}), such that $g_0(x_{10}, v) = x_{20}$ and $G_0(x_{10}, v) = x_{20}$.

Let $o(\sigma)$ denote the stochastic small-o for (σ_1, σ_2).

Proposition 2.3. Under Assumptions 2.2, 2.3, and S:

(i) The random field $M(\cdot)$ is approximately Gaussian such that:

$$M(x) = m(x) + m(x)\big[\sigma_1 U_1(x_1) + \sigma_2 U_2(x_2)\big] + o(\sigma), \tag{2.3.19}$$

for every $x \in R$.

(ii) The random field $X_1(\cdot)$ is approximately Gaussian such that:

$$X_1(z) = x_1(z) - p\Delta_x(z)\big[\sigma_1 U_1(x_1(z)) + \sigma_2 U_2(x_2(z))\big] + o(\sigma), \tag{2.3.20}$$

for every $z \in R$.

(iii) If every realization of the random field $M(\cdot)$ is continuously-differentiable, then the random field $G_0(\cdot, v)$ is well-defined. Furthermore, this random field satisfies:

$$G_0(x_1, v) = g_0(x_1, v) + h_0(x_1, v; \sigma, U_1, U_2) + o(\sigma), \tag{2.3.21}$$

for every $x_1 > 0$, where $h_0(x_1, v; \sigma, U_1, U_2)$ solves:

$$\begin{aligned} \frac{\partial h(x_1, v)}{\partial x_1} &= -\frac{\partial m}{\partial x_2}\big[x_1, g_0(x_1, v)\big] h(x_1, v) \\ &+ m(x_1, g_0(x_1, v))\big[\sigma_1 U_1(x_1) + \sigma_2 U_2(g_0(x_1, v))\big], \end{aligned} \tag{2.3.22}$$

subject to $h(x_{10}, v) = 0$.

Proof. See Appendix 2.A.7. ☐

By the linearity of the differential equation in (2.3.22):

$$h_0(x_1, v; \sigma, U_1, U_2) = \sigma_1 h_1(x_1, v) + \sigma_2 h_2(x_1, v), \tag{2.3.23}$$

for every $x_1 > 0$, where $h_j(x_1, v)$ solves:

$$\frac{\partial h_j(x_1, v)}{\partial x_1} = -\frac{\partial m}{\partial x_2}\big[x_1, g_0(x_1, v)\big] h_j(x_1, v) + m(x_1, g_0(x_1, v)) V_j(x_1, v), \tag{2.3.24}$$

subject to $h_j(x_{10}, v) = 0$, such that $V_1(x_1, v) = U_1(x_1)$ and $V_2(x_1, v) = U_2(g_0(x_1, v))$.

Small-sigma analysis is standard in physics (also known as perturbation theory), first used in Econometrics by Kadane (1971). Under the small-sigma assumption, the transformation $A(\cdot)$ is locally linearized:

$$\mathbb{E}[Z] = \text{inv}A(\mathbb{E}[X]) + o(\sigma) = \text{inv}\, x + o(\sigma) = m + o(\sigma), \tag{2.3.25}$$

in which $\text{inv}A(\cdot)$ denotes the measurable inverse of $A(\cdot)$ that transforms the demand random field $X(\cdot)$ into the inverse demand random field $Z(\cdot)$. Therefore, if σ_1 and σ_2 are sufficiently small, the expected marginal rate of substitution $\tilde{m}(\cdot)$ must be "close" to the second component of the generalized inverse of expected demand $m(\cdot)$.

Propositions 2.3(i) and 2.3(ii) describe the link between the marginal rate of substitution random field $M(\cdot)$ and the demand random field $X(\cdot)$. Note that, the knowledge of all marginal distributions of $M(\cdot)$ is not enough to know all marginal distributions of $X(\cdot)$. Proposition 2.3 provides the form of conditional heteroscedasticity on the random fields. There are two sources of heteroscedasticity for the demand random field $X(\cdot)$:

(i) a multiplier effect involving the Slutsky coefficient $\Delta_x(\cdot)$;

(ii) a deformation of space through $x(\cdot)$.

This finding is somewhat related to Brown and Walker (1989), who show that, if consumption choices are made with respect to well-behaved (but possibly distinct) utility functions, then additive disturbances on demand exhibit conditional heteroscedasticity. The result above says that, even if choices are not made with respect to well-behaved utility functions, but consumers are quasi-rational, and errors are small and Gaussian, we can model demand $X(\cdot)$ using additive disturbances with conditional heteroscedasticity.

Example 2.1 (Continued)**.** To illustrate the deformation of space, let us consider the Stone-Geary specification (with equal weights). Under this specification:

$$x(z) = \left(\frac{y}{2p}, \frac{y}{2}\right)'. \tag{2.3.26}$$

Let $\mathcal{Z}$ denote a square in the space of income and prices, defined by:

$$\mathcal{Z} = \big\{z \in R : 1 \leq p \leq 2 \text{ and } 1 \leq y \leq 2\big\}. \tag{2.3.27}$$

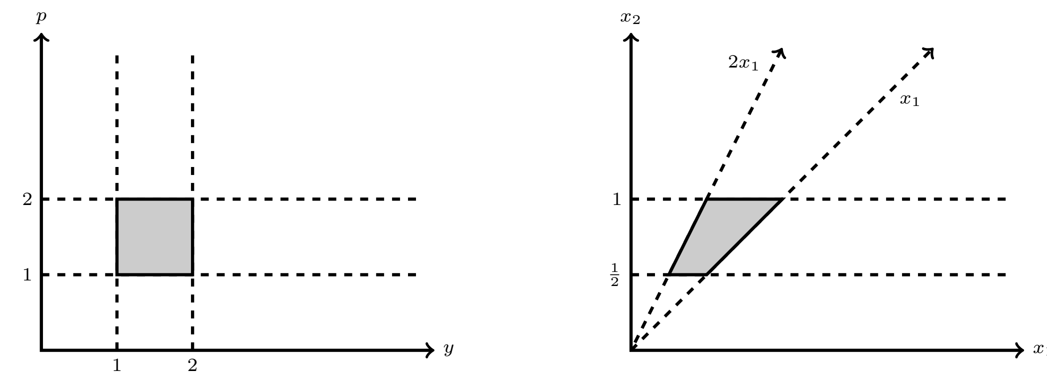

Figure 2.6. **Domain and Range in Example 2.1.** On the left, I illustrate a square in the space of income and prices from Example 2.1. On the right, I illustrate the range of demand over this square from Example 2.1.

The range $\mathcal{X}$ of the demand field $x(\cdot)$ over $\mathcal{Z}$ is not a rectangle. It has the following form:

$$\mathcal{X} = \left\{ x \in \bar{R} : \frac{x_2}{2} \leq x_1 \leq x_2 \text{ and } \frac{1}{2} \leq x_2 \leq 1 \right\}. \tag{2.3.28}$$

Figure 2.6 displays the regions, $\mathcal{Z}$ and $\mathcal{X}$. The shape of $\mathcal{X}$ follows from homotheticity. In general, demand is non-linear, resulting in a highly non-linear deformation of space. △

The following corollary is easily deduced from Proposition 2.3:

Corollary 2.1. Under Assumptions 2.2, 2.3, and S:

(i) The approximate distribution of $M(\cdot)$ has mean $m(x)$ and covariance operator:

$$C_M(x, \tilde{x}) = m(x)C(x, \tilde{x})m(\tilde{x}), \tag{2.3.29}$$

for every $x, \tilde{x} \in R$.

(ii) The approximate distribution of $X_1(\cdot)$ has mean $x_1(z)$ and covariance operator:

$$C_X(z, \tilde{z}) = p\Delta_x(z)C\big[x_1(z) - x_1(\tilde{z})\big]\tilde{p}\Delta_x(\tilde{z}), \tag{2.3.30}$$

for every $z, \tilde{z} \in R$.

Proof. This result follows immediately from Proposition 2.3. □

2.4 Identification and Estimation of the Random Fields

In this section, I introduce the assumptions on the observations, then discuss the identification and estimation of the demand random field, and the marginal rate of substitution random field. I show that the distribution of the marginal rate of substitution random field can be recovered from individual observations of consumption, whenever we have a panel with at least two observations per individual. I also show how to perform counterfactual analysis at the individual-level (with confidence bands) in the small-sigma framework.

2.4.1 Assumptions on Panel Observations

Let us consider panel data, indexed by both consumers i and dates t. In the application in Section 2.6, we observe a large number n of consumers and a small fixed number T of months. The asymptotic results are written for n tending to infinity with fixed T.

Assumption 2.4 (Latent Model)**.**

(i) Consumer i has preferences $M_i(\cdot)$, for each consumer $i = 1, \ldots, n$.

(ii) Preferences (M_i) do not depend on the month t.

(iii) Preferences (M_i) are independent and identically distributed.

Assumption 2.4 implies that the fields $(\log M_i)$ are independently drawn from the same unknown infinite-dimensional distribution and that these fields are constant over time. Assumption A4 rules out the possibility of consumers buying goods for investment or future consumption.

Assumption 2.5 (Panel Observations)**.**

(i) We observe (x_{it}, z_{it}), for each consumer $i = 1, \ldots, n$ and month $t = 1, \ldots, T$. These observations satisfy the constraint: $p_{it}x_{i1t} + x_{i2t} = y_{it}$, for each consumer $i = 1, \ldots, n$ and month $t = 1, \ldots, T$.

(ii) The quantities (x_{ijt}) are strictly positive.

(iii) Designs (z_{it}) are exogenous to consumption decisions.

Assumption 2.5(ii) lets us focus on the consumption levels, rather than the qualitative decision to consume. In the application in Section 2.6, we drop all consumers that do not consume any alcohol (see Section IV.A in Blundell et al., 2017, for a similar practice in a non-parametric analysis of gasoline demand). For this assumption to be reasonable, we need an appropriate level of aggregation. Indeed, at an extremely disaggregate level, we would have to account for *context effects* such as compromise effects, attraction effects, and similarity effects (see Huber et al., 1982, Noguchi and Stewart, 2014, Crosetto and Gaudeul, 2016, and Cataldo and Cohen, 2018, for recent papers). Context effects are not within the scope of this paper (but we can partially adjust for an effect of quality, as described in Section 2.6).

Assumption 2.5(iii) is standard, but rarely written (see Chernozhukov et al., 2020, for more on endogenous prices in scanner data). This assumption imposes a constraint on the decision process since expenditure is observed, rather than income: We can imagine consumers first deciding how much to

Table 2.1. State-space model

State Equations	$M_i(\cdot)$ and $X_i(\cdot)$, for $i = 1, \dots, n$, related by (2.3.16)
Measurement Equations	$X_{it} = X_i(Z_{it})$, for Z_{it} drawn from $\pi(\cdot)$

spend on certain groups of goods, then deciding how to spend their expenditure within each group. The first decision is based on inter-group random utility; the second decision is based on intra-group random utility. Assumption 2.5(iii) says that these decisions are independent, and that the two aggregate goods in the model are separable from all other excluded goods, but this assumption will not be tested.

From a probabilistic point of view, we can imagine the designs $Z_{it} = (Y_{it}, P_{it})$ being independently drawn from a distribution $\pi(\cdot)$ with realization z_{it}. Here, $M_i(\cdot)$ and $X_i(\cdot)$ are high-dimensional state variables, and $X_{it} = X_i(Z_{it})$ is a pair of measured quantities with realization x_{it}. Theses quantities are doubly stochastic since both $X_i(\cdot)$ and Z_{it} are stochastic. While consumption profiles $(X_{i1}, \dots, X_{iT})$, $i = 1, \dots, n$, are independent, the quantities X_{it} demanded by a given consumer i are linked since the realized preferences $M_i(\cdot)$ do not change over time. The probabilistic assumptions are summarized in Table 2.1 in terms of a state-space model. Under Assumption 2.5, we can, theoretically, derive the distribution of all observed quantities (X_{it}) given all designs (Z_{it}), and use this information to estimate the latent parameters of the model.

2.4.2 Distribution of Demand

An analysis of the demand random field $X(\cdot)$ is needed to predict the effect of a change in the distribution of designs $\pi(\cdot)$ on the distribution of demand. It will also be needed to recover the distribution of preferences in the absence of a small-sigma assumption.

Expected Demand Field

Let us first consider the expected demand field $x(\cdot)$. Under Assumptions 2.4 and 2.5:

$$x(z) \equiv \mathbb{E}[X(z)] = \mathbb{E}[X_i(z)] = \mathbb{E}[X_{it}|Z_{it} = z]. \tag{2.4.1}$$

Therefore, this field is identified, and its first component can be consistently estimated by the following Nadarya-Watson estimator:

$$\hat{x}_1(z) = \sum_{i=1}^{n} \sum_{t=1}^{T} \frac{x_{i1t} K\left(\frac{z_{it}-z}{h}\right)}{\sum_{i=1}^{n} \sum_{t=1}^{T} K\left(\frac{z_{it}-z}{h}\right)}, \tag{2.4.2}$$

where $K(\cdot)$ denotes a product kernel:

$$K\left(\frac{z_{it} - z}{h}\right) = K_1\left(\frac{z_{1it} - z_1}{h_1}\right) K_2\left(\frac{z_{2it} - z_2}{h_2}\right), \tag{2.4.3}$$

in which $K_j : \mathbb{R} \to \mathbb{R}_+$ is itself a kernel that satisfies some standard regularity conditions, and h_j is a bandwidth that tends to zero at an appropriate rate as the number of observations tends to infinity, $j = 1, 2$. After estimating the first component of $x(z)$, the budget constraint can be applied to consistently estimate the second component as:

$$\hat{x}_2(z) = y - p\hat{x}_1(z). \tag{2.4.4}$$

Consider the following regularity conditions:

Assumption 2.6.

- Conditions on the random field:
  - (i) The expected demand field $x_1(\cdot)$ is continuously-differentiable on R.
  - (ii) The random field $X_1(\cdot)$ satisfies a *weak dependence* condition.
  - (iii) The (marginal) distribution of the measured quantity X_{i1t} conditional on $Z_{it} = z$ is continuous with a strictly positive density $f_z(\cdot)$. The conditional mean $x_1(\cdot)$ and density $f_z(\cdot)$ are sufficiently regular in a neighbourhood of z.
- Condition on the designs:
  - (iv) The designs (Z_{it}) are independently and identically drawn from a continuous distribution that admits a continuous and strictly positive density $\pi(\cdot)$.
- Joint condition on the design and expected field:
  - (v) $\mathbb{E}[x_1(Z)] = \int x_1(z)\pi(z)dz$ exists.
- Conditions on the kernel:
  - (vi) $K_j(u) \geq 0$, for $u \in \mathbb{R}$ and $j = 1, 2$.
  - (vii) $\int K_j(u)du = 1$ and $\int uK_j(u)du = 0$, for $j = 1, 2$.
  - (viii) $K_j(u) = o\left(||u||^{-\xi}\right)$, with $\xi > 2$, as $u \to \infty$, for $j = 1, 2$.
- Condition on the bandwidth:
  - (ix) $h_j \to 0$ and $\sqrt{n}h_j \to \infty$ as $n \to \infty$, for $j = 1, 2$.

Assumption 2.6 focuses on the main regularity conditions (see Robinson, 2011, for additional technical assumptions needed to ensure that $x_1(\cdot)$ and $f_z(\cdot)$ are sufficiently regular), and combines the conditions needed for convergence with those needed for asymptotic normality. Assumption 2.6(ii)—a condition of "small spatial dependence" between $X_1(z)$ and $X_1(\tilde{z})$, for distinct $z, \tilde{z} \in R$—is standard. It is satisfied for the Gaussian processes in Examples 2.3 and 2.4 because they are "strong mixing" (see Robinson, 1983, for the use of a strong mixing condition). Assumption 2.6(iii) is also satisfied in these examples if

$\pi(\cdot)$ is regular, since $x_1(\cdot)$ is continuously-differentiable under Assumption 2.6(i). Assumption 2.6(iv) is introduced for expository purposes, and to account for the irregular spacing of the spatial data (Andrews, 1995). Indeed, due to the partial ordering of the bivariate index, and the economic interpretations of y and p, it would not be appropriate to assume that, for instance, $z_1 = 1, \ldots, n_1$ and $z_2 = 1, \ldots, n_2$, for some $n_1, n_2 \in \mathbb{N}$, and develop an asymptotic theory for the situation in which n_1 and n_2 tend to infinity at appropriate (relative) rates. Assumption 2.6(iv) can be relaxed in several ways. For example, we could allow for weakly dependent drawings or deterministic regular designs.[9] I do not consider such extensions to avoid writing complicated technical conditions (see Robinson, 2011). In the application, the support of $\pi(\cdot)$ is implicitly bounded, meaning that Assumption 2.6(v) is automatically satisfied as long as $x_1(\cdot)$ is continuous. The conditions on the kernels and the bandwidths are standard, and the former are satisfied by "Gaussian" kernels.

Define $\kappa \equiv \int K_1^2(u)du \int K_2^2(u)du$ and $\sigma^2(z) \equiv V[X_1(z)]$.

Proposition 2.4. Under Assumption 2.6:

$$\sqrt{nTh_1h_2}\big[\hat{x}_1(z) - x_1(z)\big] \xrightarrow{d} N\left(0, \kappa \cdot \frac{\sigma^2(z)}{\pi(z)}\right). \tag{2.4.5}$$

Furthermore, $\hat{x}_1(z)$ and $\hat{x}_1(\tilde{z})$ are asymptotically independent, whenever $z \neq \tilde{z}$.

Proof. See Robinson (2011). □

When $T = 1$, we get independently and identically distributed observations $(X_i(Z_i), Z_i)$, $i = 1, \ldots, n$, and the field aspect of demand has no effect. When $T \geq 2$, quantities demanded by a given consumer are linked. Fortunately, the Nadaraya-Watson estimator $\hat{x}(z)$ retains its consistency and asymptotic normality under weak dependence. This result follows from the local nature of the estimator.

The asymptotic variance of the Nadaraya-Watson estimator $\hat{x}(z)$ involves the deterministic fields $\sigma^2(z)$ and $\pi(z)$. These deterministic fields can be estimated by kernel approaches to obtain pointwise confidence bands for $x_1(z)$ such that:

$$\hat{x}_1(z) \pm \frac{2}{\sqrt{nTh_1h_2}} \cdot \frac{\hat{\sigma}(z)}{\hat{\pi}(z)^{1/2}}. \tag{2.4.6}$$

Uncertainty on Demand

Let us now consider how to estimate the uncertainty on the demand random field $X(\cdot)$. We focus on the description of the estimation approach. Its asymptotic properties are beyond the scope of this chapter.

(i) **Pairwise distributions of the demand random field:** The distribution of the demand random field $X(\cdot)$ is high-dimensional and difficult to estimate. However, it is rather easy to approximate

[9] Such extensions can affect the speed of convergence of the estimators, while keeping the asymptotic normality and the expression of the asymptotic variance (see Theorem 4 in Robinson, 2011).

Table 2.2. Pairs of indices for regression of Laplace transform when $T = 4$.

t	s
1	2
1	3
1	4
2	3
2	4
3	4

its pairwise distributions. Let us consider the pairwise distribution of $X_1(z)$ and $X_1(\tilde{z})$. This distribution is characterized by its pairwise Laplace transform:

$$\Psi_{z,\tilde{z}}(v,\tilde{v}) = \mathbb{E}\big[\exp\{-vX_1(z) - \tilde{v}X_1(\tilde{z})\}\big], \tag{2.4.7}$$

$$= \mathbb{E}\big[\exp\{-vX_1(Z_{it}) - \tilde{v}X_1(Z_{is})\}\,\big|Z_{it} = z, Z_{is} = \tilde{z}\big], \tag{2.4.8}$$

$$= \mathbb{E}\big[\exp\{-vX_{i1t} - \tilde{v}X_{i1s}\}\,\big|Z_{it} = z, Z_{is} = \tilde{z}\big], \tag{2.4.9}$$

where the equalities follow from the exogeneity of designs Z_{it} and the assumption that preferences (M_i) are constant across time. Notice the importance of a panel structure in this formula. Indeed, this formula cannot be used with a single cross-section. When $T \geq 2$, we can consistently estimate this (conditional) Laplace transform using a Nadaraya-Watson estimator. When $T = 2$, this procedure involves non-parametrically regressing $\exp\{-vx_{i11} - \tilde{v}x_{i12}\}$ on (z_{i1}, z_{i2}); when $T > 2$, this Laplace transform can be estimated by averaging the estimates from all pairs of dates. For example, when $T = 4$, we can non-parametrically regress $\exp\{-vx_{i1t} - \tilde{v}x_{i1s}\}$ on (z_{it}, z_{is}) using the six pairs of indices in Table 2.2 and take the average of all these estimates.

Remark 2.6. Some related literature (see Deaton and Muellbauer, 1980a, Blundell, Horowitz, and Parey, 2017, and Blundell, Kristensen, and Matzkin, 2017), assumes that the observations (X_{it}, Z_{it}), $i = 1, \ldots, n$ and $t = 1, \ldots, T$, are independent and identically distributed. When $T > 1$, this assumption requires preferences to change randomly over time, making it impossible to identify the pairwise distributions of $X(\cdot)$.

(ii) **Distribution of the demand random field:** As mentioned, we cannot expect to identify the whole distribution of the demand random field $X(\cdot)$ with only a finite number of dates T, and without any additional structure. Consequently, a Gaussian copula is introduced. This copula is sufficient for identification because it ensures that the demand random field $X(\cdot)$ is characterized by its pairwise distributions. By the arguments above, we can consistently estimate:

- The cumulative distribution function, say $F(\cdot; z)$, of $X_1(z)$;
- The transform $\Phi^{-1}F(\cdot; z)$, in which Φ is the standard normal cumulative distribution function;

- The covariance operator of this transform:

$$C^*(z,\tilde{z}) = \operatorname{cov}\big[\Phi^{-1}F(\cdot;z), \Phi^{-1}F(\cdot;\tilde{z})\big]. \tag{2.4.10}$$

Therefore, we can approximate the distribution of the demand random field $X(\cdot)$ by a distribution with a Gaussian copula field. This approximation is such that the transform $\Phi^{-1}F(\cdot;z)$, z varying, is a Gaussian random field with zero-mean and covariance operator $C^*(\cdot)$. Gaussian copulas make it easy to simulate $X(\cdot)$. This estimator is consistent when the distribution of $\Phi^{-1}F(\cdot;z)$ is Gaussian, as under a small-sigma assumption.

2.4.3 Distribution of the Marginal Rate of Substitution

I consider two estimation approaches for the distribution of the marginal rate of substitution random field $M(\cdot)$, depending on whether errors are small or large. In the small-sigma framework (see Section 2.3.4), the impact of the "causal direction" can be neglected, allowing for the direct estimation of this distribution. Else, the estimation of this distribution is indirect.

In the small-sigma framework, the random fields, $M(\cdot)$ and $X(\cdot)$, are approximately Gaussian, and the estimators of $\mu(\cdot)$, $C(\cdot)$, and (σ_1, σ_2) are consistent as n tends to infinity with fixed T, whenever $T \geq 2$, implying that the distribution of the marginal rate of substitution random field $M(\cdot)$ is identified. A similar result holds in the general framework, whenever the approximation of the demand random field $X(\cdot)$ in Section 2.4.2 is consistent.

The observational equivalence result in Hausman and Newey (2016a), implying that only one-dimensional distributions of heterogeneity can be identified, does not apply because we have panel observations, allowing us to consider more than just the marginal distributions of the demand random field.

Small-Sigma Framework

The relationship in Assumption 2.3 implies that, when errors are small, the functional parameter $\mu(\cdot)$ can be estimated by non-parametrically regressing $\log p_{it}$ on x_{it} using the Nadaraya-Watson approach. The consistency of this estimator holds under standard regularity conditions, as long as σ_1 and σ_2 tend to zero at appropriate rates.[10] It is left to estimate the covariance operator $C(\cdot)$ and the parameters (σ_1, σ_2). This can be done by parameterizing $C_j(\cdot)$, and choosing the parameters that minimize the sum of squared residuals.

Let us introduce the following assumption:

Assumption 2.7. The operator $C_j(\cdot)$ is parameterized by $k_j \in \mathbb{R}^{L_j}$, for each $j = 1, 2$.

[10] If σ_1 and σ_2 tend to zero very quickly, then we will also obtain a standard asymptotic result for this estimator.

Under Assumption 2.7, the distribution of $M(\cdot)$ is characterized by the functional parameter $\mu(\cdot)$, and the scalar parameters $\theta = (k_1, k_2, \sigma_1, \sigma_2)$. Thus, after estimating $\mu(\cdot)$, we can compute the residuals:

$$\hat{\varepsilon}_{it} = \log p_{it} - \hat{\mu}(x_{it}), \tag{2.4.11}$$

and choose θ to minimize:

$$\sum_{i=1}^{n} \sum_{t \leq s} \left[\hat{\varepsilon}_{it}\hat{\varepsilon}_{is} - C_\theta(x_{it} - x_{is})\right]^2, \tag{2.4.12}$$

in which $C_\theta(\cdot)$ denotes the covariance operator for $\log M(\cdot)$ given θ. This approach is consistent when there exists a unique mapping between k_j and $C_j(\cdot)$ over the range of demand. This approach is only possible when the causal direction can be neglected. Assumption 2.7 makes estimation simple, but it is not necessary (see Section 2.4.3). The asymptotics of the estimators above are straightforward and omitted for brevity.

General Framework

The non-parametric estimation of the latent model is more difficult if the uncertainty on the marginal rate of substitution random field $M(\cdot)$ is large. Once again, I focus on the description of the approach.

A non-parametric estimate of the distribution of the marginal rate of substitution random field $M(\cdot)$ can be recovered by simulating from the estimated distribution of the demand random field $X(\cdot)$. Consider the following steps:

Step 1: Estimate the marginal and pairwise distributions of $X_1(\cdot)$, and then approximate the distribution of $X_1(\cdot)$ using a Gaussian copula (see Section 2.4.2);

Step 2: Simulate the demand field: $\hat{X}_1^s = \{\hat{X}_1^s(z),\, z \text{ varying}\}$, for $s = 1, \ldots, S$ (see Appendix 2.D);

Step 3: Use (2.2.12) to deduce the marginal rate of substitution $\hat{M}^s = \{\hat{M}^s(x),\, x\text{varying}\}$, for $s = 1, \ldots, S$, and apply the log-transform to get $\log \hat{M}^s = \{\log \hat{M}^s(x),\, x \text{ varying}\}$, for $s = 1, \ldots, S$;

Step 4: Estimate the functional parameter $\mu(\cdot)$ using:

$$\hat{\mu}^*(x) = \frac{1}{S} \sum_{s=1}^{S} \log \hat{M}^s(x), \tag{2.4.13}$$

then approximate σ_j and $C_j(\cdot)$ with sample counterparts, computed by simulation, for $j = 1, 2$.

Testing the Small-Sigma Assumption

The arguments above elicit a question: Can we neglect the causal direction? There are several ways to obtain an answer to this question. For example, we can compare the second component, say $\hat{m}(\cdot)$, of the inverse of the estimate $\hat{x}(\cdot)$ of the expected demand field $x(\cdot)$ with the exponential transform of the estimate $\hat{\mu}(\cdot)$ of the functional parameter $\mu(\cdot)$, obtained in the small-sigma framework. Alternatively, we can compare the small-sigma estimate $\hat{\mu}(\cdot)$ with the general estimate $\hat{\mu}^*(\cdot)$. In either case, we can neglect the causal direction if, and only if, the difference between these objects is small.

2.4.4 Filtering and Counterfactual Analysis

Once the random field model is estimated using one of the methods above, this model can be used for filtering (i) the unobserved part of the individual demand $X_{1i}(z)$, (ii) the unobserved individual marginal rate of substitution $M_i(x)$, (iii) the unobserved individual errors $M_i(x) - m(x)$ and their decomposition in the direction of each good, and (iv) the unobserved individual indifference curves $g(x, v)$. These filtered curves have to be provided with their prediction bands. The filtering step requires a space (or space-time) type of a Kalman-filter (see Wikle and Cressie, 1999). In order to apply the result in this section, we first need to estimate expected demand and the distribution of marginal rate of substitution (see Section 2.4.3).

Filtering is easy under the small-sigma assumption because all of the random fields are locally Gaussian (see Proposition 2.3). Let us consider the prediction of the demand of consumer i at some counterfactual value z. For this individual, we observe $X_{1i}(z_{it})$, for each $t = 1, \dots T$. By the independence restriction in Assumption 2.5, $X_{1i}(z_{it})$, $t = 1, \dots, T$, is a sufficient statistic for predicting $X_{1i}(z)$, at any counterfactual value z, and by Proposition 2.3(ii), the random field $X_{1i}(\cdot)$ is locally Gaussian.

As a consequence, we obtain the following result:

Proposition 2.5. Under Assumptions 2.2 to 2.6, and S, the conditional distribution of $X_{1i}(z)$ given $X_{1i}(z_{it})$, $t = 1, \dots, T$, is Gaussian with mean:

$$x_1(z) + \gamma' \Omega^{-1} \big[X_{1i}(z_{i1}) - x_1(z_{i1}), \dots, X_{1i}(z_{iT}) - x_1(z_{iT}) \big]', \tag{2.4.14}$$

and variance $V(X_{1i}(z)) - \gamma' \Omega^{-1} \gamma$ such that:

$$\begin{aligned} \Omega &= V([X_{1i}(z_{i1}), \dots, X_{1i}(z_{iT})]') \\ \gamma &= \mathrm{cov}(X_{1i}(z), [X_{1i}(z_{i1}), \dots, X_{1i}(z_{iT})]'). \end{aligned} \tag{2.4.15}$$

Proof. See Wikle and Cressie (1999). □

The elements of the matrix Ω and vector γ have closed-form expressions in terms of the covariance operators of $U_1(\cdot)$ and $U_2(\cdot)$. For instance:

$$\begin{aligned} \Omega_{t,s} &= \mathrm{cov}(X_{1i}(z_{it}), X_{1i}(z_{is})) \\ &= p_{it} p_{is} \Delta_x(z_{it}) \Delta_x(z_{is}) \big[\sigma_1^2 C_1(x_1(z_{it}), x_1(z_{is})) + \sigma_2^2 C_2(x_2(z_{it}), x_2(z_{is})) \big]. \end{aligned} \tag{2.4.16}$$

The distributional result in Proposition 2.5 can be directly used to construct prediction intervals for counterfactual $X_{1i}(z)$ after replacing the scalar and functional parameters by their estimates. Clearly, the approach in Proposition 2.5 can be extended to (i) the joint prediction of several counterfactual demands $X_{1i}(z_1), \ldots, X_{1i}(z_J)$, (ii) the construction of an impulse response function, $\delta \mapsto X_{1i}(z_{iT}+\delta z)-X_{1i}(z_{iT})$, (iii) the prediction of a counterfactual marginal rate of substitution $M_i(\cdot)$ using the joint expansions in Proposition 2.3, or (iv) aggregate counterfactual analysis. When errors are not small, counterfactual analysis can be performed by means of numerical methods.

2.5 Representative Consumer

The existence of a rational "representative consumer" is non-trivial. Indeed, a rational representative consumer can fail to exist, even if every consumer is rational, and can exist, even if every consumer violates rationality. An analysis of the representative consumer can be useful for constructing a simple and coherent notion of consumer welfare, even if every consumer violates rationality (see Section 3.4 in Blundell et al., 2003).

2.5.1 Definition

I define the representative consumer as the consumer whose demand field is the expected demand field (see Gorman, 1953, Muellbauer, 1976, Grandmont, 1992, Hildenbrand, 1994, and Blundell et al., 2003, for more information on this topic).[11]

This definition of a representative consumer differs from the definition in Gorman (1953) and Muellbauer (1976). In these papers, prices are constant across consumers, and the quantity demanded by the representative consumer is defined to be the sum:

$$\bar{x}_{jt} \equiv \sum_{i=1}^{n} x_{ijt}. \tag{2.5.1}$$

In the current framework, this definition is not appropriate because prices vary significantly across consumers in scanner data (see Section 2.6).

If the Slutsky coefficient $\Delta_x(\cdot)$ for the expected demand field $x(\cdot)$ is strictly negative, then this field $x(\cdot)$ is a *bona fide* demand field, and it can be inverted to obtain the marginal rate of substitution $m(\cdot)$ of a rational representative consumer.

In general, the marginal rate of substitution of the representative consumer $m(\cdot)$, obtained by inverting the expected demand field (when possible), does not equal the expectation of the marginal rate of

[11] Gorman (1953) and Muellbauer (1976) assume that individuals are rational. In our framework, consumers are quasi-rational. Therefore, our framework is closer to Grandmont (1992), Hildenbrand (1994), and Blundell et al. (2003).

substitution field $\tilde{m}(\cdot)$. Because $A(\cdot)$ is non-linear:

$$\begin{aligned}\mathbb{E}[Z] = \mathbb{E}[\text{inv}A(X)] \neq \text{inv}A(\mathbb{E}[X]) = \text{inv}A(x),\\ x = \mathbb{E}[X] = \mathbb{E}[A(Z)] \neq A(\mathbb{E}[Z]).\end{aligned} \tag{2.5.2}$$

However, when σ_1 and σ_2 are small, these two functions are "close" (see Section 2.3.4).

Consider an example: f

Example 2.2 (Continued). Under the Stone-Geary specification in Example 2.2, defined by $u(x) = x_1^{\alpha} x_2^{1-\alpha}$, for some $\alpha \in (0,1)$ and any $x \in \bar{R}$:

$$\log M(x) = \log(x_2) - \log(x_1) + \lambda_m + \sigma^2 U, \tag{2.5.3}$$

where U is standard normal. By Proposition 2.2(ii), the mean of $M(\cdot)$ equals:

$$\tilde{m}(x) = \frac{x_2}{x_1} \cdot e^{\lambda_m + \frac{\sigma^2}{2}}, \tag{2.5.4}$$

for each $x \in R$. For each realization of $M(\cdot)$, demand $X_1(\cdot)$ solves:

$$\frac{y - pX_1(z)}{X_1(z)} \cdot e^{\lambda_m + \sigma^2 U} - p = 0. \tag{2.5.5}$$

This equation has a unique solution:

$$X_1(z) = \left(\frac{e^{\lambda_m + \sigma^2 U}}{1 + e^{\lambda_m + \sigma^2 U}} \right) \frac{y}{p}. \tag{2.5.6}$$

Now, suppose $\lambda_m = 0$. Under this assumption, the term in parentheses has a (standard) logit-normal distribution. Since U has zero-mean, the probability density function of this term is symmetric on $(0,1)$, with a mean equal to $1/2$. Therefore, we have:

$$x_1(z) = \frac{y}{2p}, \tag{2.5.7}$$

for each $z \in R$, so that the expected demand field $x(\cdot)$ coincides with the deterministic demand field for the Stone-Geary specification with equal weights in Example 2.1. Let $m(\cdot)$ denote the marginal rate of substitution field of this utility function. The fields, $\tilde{m}(\cdot)$ and $m(\cdot)$, coincide if the preferences are deterministic such that $\sigma^2 = 0$. Otherwise, there is a convexity bias:

$$m(x) = \frac{x_2}{x_1} \neq \frac{x_2}{x_1} \cdot e^{\frac{\sigma^2}{2}} = \tilde{m}(x). \tag{2.5.8}$$

△

2.5.2 Preferences of the Representative Consumer

In the small-sigma framework, the preferences of the representative consumer can be recovered by applying the exponential transform to $\hat{\mu}(\cdot)$ because $\exp \mu(\cdot) \simeq m(\cdot)$. In the general framework, the expected demand field $x(\cdot)$ needs to be inverted. I focus on the general estimation approach since the small-sigma approach is straightforward. Let $\hat{m}(\cdot)$ denote the second component of the generalized inverse of the estimate $\hat{x}(\cdot)$.

Define $\sigma_z^2(x) \equiv \sigma^2(z(x))$ and $\pi_z(x) \equiv \pi(z(x))$.

Corollary 2.2. Under Assumptions 2.2 to 2.6, if $\Delta_x(z) < 0$, for any $z \in \operatorname{inv} x(z)$, then:

$$\sqrt{nTh_1h_2}\big[\hat{m}(x) - m(x)\big] \xrightarrow{d} N\left(0, \kappa \cdot \frac{\sigma_z^2(x)}{\pi_z(x)} \cdot \Delta_x(z(x))^{-2}\right). \tag{2.5.9}$$

Proof. See Appendix 2.A.8. □

Corollary 2.2 provides the asymptotic properties of the general estimator $\hat{m}(\cdot)$ for the marginal rate of substitution $m(\cdot)$ of the representative consumer, obtained by inverting the estimate $\hat{x}(\cdot)$ of the expected demand field $x(\cdot)$. The negativity of the Slutsky coefficient $\Delta_x(z)$ guarantees the invertibility of $\hat{x}(\cdot)$ in an open neighbourhood of z.

2.5.3 Slutsky Coefficient

If we want to estimate the Slutsky coefficient $\Delta_x(\cdot)$, either for counterfactual analysis (see Section 2.4.4), estimating the accuracy of $\hat{m}(\cdot)$ (see Section 2.5.2), or testing the integrability of the expected demand field $x(\cdot)$ (see Section 2.5.4), then we need estimators for the partial derivatives of $x_1(\cdot)$, and the joint asymptotic distribution of these estimators after an appropriate normalization. Historically, first-order partial derivatives have been estimated by estimating the function $\hat{x}_1(\cdot)$, then evaluating $\partial\hat{x}_1(z)/\partial z_i$ (Vinod and Ullah, 1987; Ullah, 1988), or $\big[\hat{x}_1(z+h) - \hat{x}_1(z)\big]/h$, for a small value of $h > 0$ (Rilstone, 1985).[12] These procedures are not appropriate in the current framework because we will eventually require the *joint* asymptotic distribution of these estimators. Therefore, I propose the application of a *local quadratic fit*.

The Nadaraya-Watson estimator $\hat{x}_1(\cdot)$ is the solution to the following weighted least squares optimization problem:

$$\begin{aligned}\hat{x}_1(z) = \operatorname*{argmin}_a \sum_{i=1}^{n}\sum_{t=1}^{T}\Bigg[&\frac{1}{h_1}K_1\left(\frac{z_{1it}-z_1}{h_1}\right) \\ &\cdot\frac{1}{h_2}K_2\left(\frac{z_{2it}-z_2}{h_2}\right)(x_{1it}-a)^2\Bigg],\end{aligned} \tag{2.5.10}$$

where $x_1(z)$ is locally approximated by some constant. This interpretation can be extended by considering a local approximation by a polynomial of degree 2 in the components of z, that is, a local quadratic

[12] A similar approach has been developed in Gasser and Müller (1984) for deterministic regressors.

fit. This approach was initially introduced by Cleveland (1979). Then, the objective function becomes:

$$\mathcal{L}(z|a,b,c) = \sum_{i=1}^{n}\sum_{t=1}^{T}\left[\frac{1}{h_1}K_1\left(\frac{z_{1it}-z_1}{h_1}\right)\frac{1}{h_2}K_2\left(\frac{z_{2it}-z_2}{h_2}\right)\right.$$
$$\left.\cdot\left\{x_{it}-\left[a+b'\begin{pmatrix} z_{1it}-z_1 \\ z_{2it}-z_2\end{pmatrix}+\frac{1}{2}\begin{pmatrix} z_{1it}-z_1 \\ z_{2it}-z_2\end{pmatrix}' c\begin{pmatrix} z_{1it}-z_1 \\ z_{2it}-z_2\end{pmatrix}\right]\right\}^2\right], \tag{2.5.11}$$

where a is a scalar, b is a vector, and c is a positive symmetric matrix. The solutions, $\hat{a}(z)$, $\hat{b}(z)$, and $\hat{c}(z)$, of the resulting minimization problem estimate $x_1(z)$, $\partial x_1(z)/\partial z$, and $\partial^2 x_1(z)/\partial z\partial z'$, respectively. These non-parametric estimators are consistent estimators of their theoretical counterparts, with rates of convergence that depend on the degree of differentiation (see Section 2.5.2, and Fan and Gijbels, 1992, Ruppert and Wand, 1994, and Masry, 1996). After estimating these partial derivatives, the Slutsky coefficient $\Delta_x(\cdot)$ can be estimated by:

$$\hat{\Delta}_x(z) \equiv \hat{b}_2(z) + \hat{a}(z)\hat{b}_1(z). \tag{2.5.12}$$

In the small-sigma framework, a similar approach can be used to estimate the coefficient $\Delta_m(\cdot)$. Otherwise, the estimate $\hat{\Delta}_x(z)$ can be plugged into the expression in (2.2.11).

The application of a local quadratic fit has two advantages:

(i) This method is easy to implement.

(ii) It is known that the estimator $\hat{a}(z)$ differs from the Nadaraya-Watson estimator. It is equivalent to the Nadaraya-Watson estimation at order $(nTh_1h_2)^{-1/2}$, but is more efficient if we consider the higher order terms. Hence, the local quadratic fit is more informative than the local constant fit.

Consider the following result:

Proposition 2.6. Under Assumptions 2.6 and C:

$$\sqrt{nTh_1h_2}\cdot D(h)\cdot\left[\begin{pmatrix}\hat{a}(z)-x_1(z)\\ \hat{b}(z)-\frac{\partial x_1(z)}{\partial z}\\ \text{vech }\hat{c}(z)-\text{vech }\frac{\partial^2 x_1(z)}{\partial z\partial z'}\end{pmatrix}-B(h,z)\right]\xrightarrow{d} N(0,\Sigma(z)), \tag{2.5.13}$$

where $D(h) = \text{diag}(1, h_1h_2, h_1^2h_2^2)$ and vech$(\cdot)$ denotes the operator stacking the distinct elements of a symmetric matrix, $B(h,z)$ denotes asymptotic bias, and $\Sigma(z)$ denotes asymptotic variance.

Proof. See Appendix 2.C and Lu (1996) for additional details. □

When appropriately normalized and bias adjusted, the local quadratic fit produces asymptotically normal estimators. These estimators are asymptotically independent when the degrees of differentiation differ—equivalently, when the rates of convergence differ (Lu, 1996). Proposition 2.6 needs an extra assumption, described in Appendix 2.C. Standard estimation methods for weighted least squares can be used to estimate the variance-covariance matrix $\Sigma(z)$.

By applying an appropriate version of the delta-method, we can also deduce the asymptotic behaviour of the plug-in estimator $\hat{\Delta}_x(z)$ of the Slutsky coefficient $\Delta_x(z)$, as defined in (2.5.12). The variability of the estimator for the expected demand field $x(\cdot)$ can be neglected since it converges faster than the estimators of its partial derivatives.

Corollary 2.3. Under Assumptions 2.2 to 2.6, and C:

$$\sqrt{nTh_1^3h_2^3}\left[\hat{\Delta}_x(z) - \Delta_x(z)\right] \xrightarrow{d} N\big(0, \Sigma_{33}(z) + x_1(z)^2\Sigma_{22}(z) + 2x_1(z)\Sigma_{32}(z)\big), \tag{2.5.14}$$

where $\Sigma_{jk}(z)$ denotes the $(j,k)^{th}$-entry of the (asymptotic) variance-covariance matrix $\Sigma(z)$. Furthermore, $\hat{\Delta}_x(z)$ and $\hat{\Delta}_x(\tilde{z})$ are asymptotically independent, whenever $z \neq \tilde{z}$.

Proof. See Appendix 2.C. □

As expected, the plug-in estimator $\hat{\Delta}_x(z)$ of the Slutsky coefficient converges at a non-parametric rate since it depends on the partial derivatives of $x_1(\cdot)$. A similar asymptotic result can be obtained for a direct estimator $\hat{\Delta}_m(\cdot)$ of the coefficient $\Delta_m(\cdot)$.

2.5.4 Testing Integrability

We are now in a position to check the integrability of the expected demand field $x(\cdot)$. I consider a null hypothesis of the form: $\mathrm{H}_{0,x} = \{\Delta_x(z) < 0, \ \forall z \in \mathcal{Z}\}$.[13] Therefore, we need to test an uncountable set of inequalities. This hypothesis is used to test an analogue of the moment inequalities in the revealed preference literature with a small number of designs (Samuelson, 1938; Houthakker, 1950; Afriat, 1967).

Some sub-hypotheses can be tested at parametric rates, while others can only be checked at non-parametric rates. Currently, it is not known how to simply put together test statistics converging at different rates. To understand this difficulty, let us discuss some standard approaches:

(i) We can construct pointwise one-sided confidence bands for the Slutsky coefficient $\Delta_x(\cdot)$ using the asymptotic distribution in Corollary 2.3, and check whether these bands are above zero. If they contain zero, at some value of $z \in \mathcal{Z}$, then optimizing behaviour cannot be accepted at z. This approach can reveal the subset of designs $\mathcal{Z}^* \subseteq \mathcal{Z}$ on which integrability is accepted and its complement on which it is not (see Dette et al., 2016, for an example of this type of approach, and Section 2.6.4). This approach is easy to implement and informative on $\mathcal{Z}^*$, but it is pointwise (making it hard to control for type I error), and not very powerful in the current framework because $\hat{\Delta}_x(\cdot)$ converges at a non-parametric rate.

[13] Alternatively, we could consider a null hypothesis of the form: $\mathrm{H}_{0,m} = \{\Delta_m(x) > 0, \ \forall x \in \mathcal{X}\}$.

(ii) Alternatively, we can consider the hypothesis:

$$\mathrm{H}'_{0,x} = \Big\{ \sup_{z\in\mathcal{Z}} \Delta_x(z) < 0 \Big\}. \tag{2.5.15}$$

This hypothesis leads to a test statistic with the form:

$$\xi_x = \sup_{z\in\mathcal{Z}} \hat{\Delta}_x(z). \tag{2.5.16}$$

This approach requires the derivation of the asymptotic distribution of ξ_x under the least favourable distribution satisfying the null hypothesis (often by simulation). This approach has four drawbacks: First, it can be difficult to precisely estimate the distribution of ξ_x because the joint distribution of $\hat{\Delta}_x(z)$, z varying, is unknown. Second, if the null hypothesis is rejected, then we have no information on the subset $\mathcal{Z}^* \subseteq \mathcal{Z}$ on which integrability is satisfied. Third, if $\mathrm{H}'_{0,x}$ is accepted, then $\mathrm{H}_{0,x}$ is also accepted, but rejecting $\mathrm{H}'_{0,x}$ does not imply that $\mathrm{H}_{0,x}$ can be rejected. For example, if $\mathcal{Z} = R$ and the utility function is Stone-Geary with equal weights (as in Example 2.1), then ξ_x becomes:

$$\xi_x = \sup_{z\in\mathcal{Z}} -\frac{y}{4p^2} = 0, \tag{2.5.17}$$

even though $\Delta_x(z) < 0$ on $\mathcal{Z}$. Finally, this test also has a problem with power.

(iii) A third approach consists of testing weak forms of integrability. The hypothesis $\mathrm{H}_{0,x}$ is equivalent to the following hypothesis:

$$\Big\{ \int_{\mathcal{Z}} \Delta_x(z) f(z) dz < 0, \text{ for any density } f \text{ with bounded support} \Big\}. \tag{2.5.18}$$

Therefore, we can take a subset $\mathcal{F} = \{f_1(z), \ldots, f_J(z)\}$ of densities and consider:

$$\mathrm{H}''_{0,x}(\mathcal{F}) = \Big\{ \int_{\mathcal{Z}} \Delta_x(z) f_j(z) dz < 0, \forall j = 1, \ldots, J \Big\}. \tag{2.5.19}$$

In other words, we can consider a sub-hypothesis consisting of moment inequality restrictions. If $\mathrm{H}''_{0,x}(\mathcal{F})$ is rejected, then $\mathrm{H}_{0,x}$ is also rejected. Note that, the asymptotic distribution derived in Corollary 2.3 cannot be used directly. Indeed, by averaging the function $\hat{\Delta}_x(\cdot)$ with an appropriate choice for $f_j(\cdot)$, we will obtain a parametric rate of convergence. A similar approach was applied in Lewbel (1995) to test the symmetry of the Slutsky matrix (a necessary condition automatically satisfied in our two good framework). This approach is conservative because it considers a finite set of functions $f_j(\cdot)$ (which can be extended to a larger family of functions if desired), but much less conservative than the pointwise approach in Dette et al. (2016) because of the parametric rate of convergence. A second advantage of this approach is to avoid differentiating the demand field $x(\cdot)$ under an appropriate choice of $\mathcal{F}$. To illustrate, let us consider one function $f_j(\cdot)$ that is

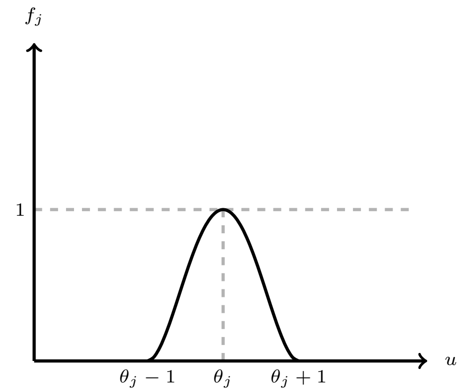

Figure 2.7. The drifted square of an Epanechnikov kernel.

continuously-differentiable with compact support satisfying separability:

$$f_j(z) = f_{1j}(y)f_{2j}(p), \tag{2.5.20}$$

where $f_{1j}(\cdot)$ is non-negative and continuously-differentiable with support $[\underline{y}, \bar{y}]$ and $f_{2j}(\cdot)$ is non-negative and continuously-differentiable with support $[\underline{p}, \bar{p}]$. An example of such a function can be deduced from the square of an Epanechnikov kernel (sometimes known as a biweight kernel) by drifting the following function:

$$f(u) = \left[1 - u^2\right]^2 \mathbb{1}_{|u|<1}. \tag{2.5.21}$$

This function has support $[-1, 1]$ and its derivatives at ± 1 are zero. The resulting drifted function $f_{ij}(u) = f(u - \theta_{ij})$, $u \in \mathbb{R}$, is non-negative and continuously-differentiable with support $[\theta_{ij} - 1, \theta_{ij} + 1]$ (see Figure 2.7). If we consider the expression for $\Delta_x(z)$ in (2.2.10) and the product function in (2.5.20), then we obtain:

$$\begin{aligned} \int \Delta_x(z) f_j(z) dz &= \int \frac{\partial x_1(z)}{\partial p} f_j(z) dz + \int x_1(z) \frac{\partial x_1(z)}{\partial y} f_j(z) dz \\ &= \iint \frac{\partial x_1(z)}{\partial p} f_{1j}(y) f_{2j}(p) dy dp + \frac{1}{2} \iint \frac{\partial x_1^2(z)}{\partial y} f_{1j}(y) f_{2j}(p) dy dp. \end{aligned} \tag{2.5.22}$$

Integrating by parts yields:

$$-\iint x_1(z) f_{1j}(y) \frac{d f_{2j}(p)}{dp} dy dp - \frac{1}{2} \iint x_1(z)^2 \frac{d f_{1j}(y)}{dy} f_{2j}(p) dy dp. \tag{2.5.23}$$

Therefore, instead of estimating this quantity by first estimating the partial derivatives of $x_1(\cdot)$, then using these derivatives to construct $\int \hat{\Delta}_x(z) f_j(z) dz$, we can simply estimate the expected demand field $x(\cdot)$ and plug its first component into (2.5.23). Finally, consider the difference between this estimator and its limit:

$$\begin{aligned}\int \left[\hat{\Delta}_x(z) - \Delta_x(z)\right] f_j(z)dz = &- \iint \left[\hat{x}_1(z) - x_1(z)\right] f_{1j}(y)\frac{df_{2j}(p)}{dp}dydp \\ &- \frac{1}{2}\iint \left[\hat{x}_1(z)^2 - x_1(z)^2\right]\frac{df_{1j}(y)}{dy}f_{2j}(p)dydp.\end{aligned} \tag{2.5.24}$$

This quantity approximately equals:

$$-\iint \left[\hat{x}_1(z) - x_1(z)\right]\left[f_{1j}(y)\frac{df_{2j}(p)}{dp} + x_1(z)\frac{df_{1j}(y)}{dy}f_{2j}(p)\right]dydp. \tag{2.5.25}$$

This quantity is a linear functional of the difference $\hat{x}_1(z) - x_1(z)$. It is known that, after applying a partition of unity (to pass from local to global analysis) and standardizing by the parametric rate $1/\sqrt{nT}$, this integral is asymptotically normal, under some appropriate conditions on $x_1(\cdot)$, $f_{1j}(\cdot)$, and $f_{2j}(\cdot)$. This integral depends on $f_j(\cdot)$, defining a Gaussian process with respect to this function, as long as the moments exist (see Theorem 4 in Zinde-Walsh, 2018).

Define $f(z) = \left[f_1(z), \ldots, f_J(z)\right]'$.

Proposition 2.7. Under Assumptions 2.2 to 2.6:

$$\sqrt{nT}\left[\int_{\mathcal{Z}} \hat{\Delta}_x(z)f(z)dz - \int_{\mathcal{Z}} \Delta_x(z)f(z)dz\right] \xrightarrow{d} N(0, \Sigma^*), \tag{2.5.26}$$

where Σ^* is a variance-covariance matrix whose $(j,k)^{th}$-entry is equal to:

$$w_{jk} = \int_{\mathcal{Z}} \frac{\sigma^2(z)}{\pi(z)}\Psi_j(z)\Psi_k(z)dz, \tag{2.5.27}$$

in which $\Psi_j(z)$ has the following form:

$$\Psi_j(z) = f_{1j}(y)\frac{df_{2j}(p)}{dp} + x_1(z)\frac{df_{1j}(y)}{dy}f_{2j}(p). \tag{2.5.28}$$

Proof. See Theorem 4 in Zinde-Walsh (2018). □

The asymptotic variance-covariance matrix can be easily estimated by replacing $x_1(z)$, $\sigma^2(z)$, and $\pi(z)$ with their sample counterparts: $\hat{x}_1(z)$, $\hat{\sigma}^2(z)$, and $\hat{\pi}(z)$. The procedure to test a finite number of inequality contraints is well-established (see Chapter 21 in Gouriéroux and Monfort, 1995, or Andrews and Shi, 2013) and can be extended to an infinite number of inequality constraints if desired. Theorem 2.7 implies that the asymptotic distribution of the integral in (2.5.26) does not depend on the kernel $K(\cdot)$.

2.6 Application to the Consumption of Alcohol

In this section, the approaches presented in the previous section are applied to an analysis of preferences for different types of alcohol. I focus on estimation under the small-sigma assumption.

2.6.1 The Data

For this analysis, I use the Nielsen Homescan Consumer Panel (NHCP). Nielsen provides each household in the NHCP with a barcode scanner. Households are asked to scan every packaged commodity that they purchase. Prices are entered by the household or linked to retailer data by Nielsen. Households are financially compensated for participation through benefits and lotteries, and self-select into participation. This feature could create a self-selectivity bias that is neglected in the analysis below (see Appendix 2.B for a discussion of the representativeness of the sample, and Chernozhukov et al., 2020, for an alternative use of this dataset).

I focus on the consumption of alcoholic beverages. Beverages are grouped by type, as described in Section 2.F. Good 1 contains beers and ciders.[14] Good 2 contains wines and liquors. I omit non-alcoholic beers, ciders, and wines. There are 30,635 types of beers and ciders, and 108,439 types of wines and liquors, implying a total of 139,074 types of beverages.

All units are converted to litres, taking into account that some products are in packs of, say, six, or twenty-four.[15] For example, if a household buys two packs of six bottles of beer and each bottle contains 355 millilitres of beer, then it buys 4.26 litres of beer. The NHCP does not contain information about alcohol by volume (ABV), but we could multiply the volume of the beverage in litres by the average ABV for the type of drink (e.g., 4.5% for beer and cider, 11.6% for wine, and 37% for liquor) to diminish the endogeneity of prices that can follow from the joint decision of quality and quantity (see Chernozhukov et al., 2020, for a test of endogeneity, the discussion of context effects in Section 2.4.1, and Chapter 3 for an example of this practice). I do not make this adjustment, for simplicity.

I restrict our sample to purchases made in August to November of 2016. These months are consecutive, and avoid several holidays with special alcohol consumption such as Independence Day, Christmas Day, and New Year's Eve.[16] This window length also diminishes the impact of changing tastes and product availability.

I aggregate each household's purchases by month and restrict our sample to households with strictly positive expenditure in each of the two goods in each month (see Section IV.A in Blundell et al., 2017, for a similar assumption in an application to gasoline demand). This procedure leaves us a total of 773 households and 3,092 observations. The sample characteristics are provided in Appendix 2.B.

For each household and month, the price of an aggregate group is constructed by dividing total

[14] The NHCP classifies ciders as wine, by default. I reclassify these beverages using UPC product descriptions because most ciders have a low Alcohol By Volume (ABV).

[15] Some units were manually coded—for example, a five litre Heineken mini-keg.

[16] Although, admittedly, this window still includes Thanksgiving Day.

Table 2.3. Summary of expenditures, prices, and consumptions including the mean, standard deviation, ratio of the standard deviation to the mean, and quantiles. Units are dollars per litre.

				Quantiles				
Variable	Mean	Std. Dev.	Ratio	0%	25%	50%	75%	100%
$\tilde{y}$	146.26	136.85	0.93	6.23	66.97	111.11	184.05	2767.76
$\tilde{p}_1$	2.82	1.32	0.47	0.63	1.97	2.51	3.40	21.98
$\tilde{p}_2$	10.13	7.27	0.71	0.70	5.33	8.29	12.59	124.95
y	19.56	22.13	1.13	0.33	7.19	13.26	24.54	354.05
p	0.38	0.28	0.74	0.02	0.20	0.31	0.49	6.17
x_1	25.51	42.20	1.65	0.24	8.51	16.77	31.93	1080.81
x_2	10.94	15.07	1.37	0.05	3.00	6.53	13.34	225.75

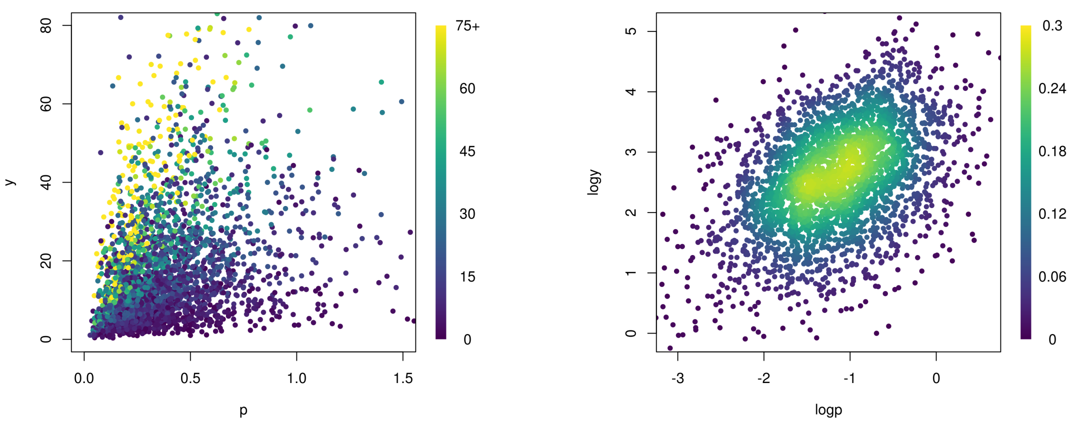

Figure 2.8. Sample distribution of normalized designs. On the left, colour indicates consumption of good 1. On the right, colour indicates a bivariate kernel density estimate.

expenditure in the group (after subtracting the value of coupons) by the litres purchased in the group, as described in Appendix 2.F. Then, I normalize by the aggregate price of wine and liquor. To illustrate the heterogeneity in expenditure and prices, summary statistics for these variables, pooled across households and months, are provided in Table 2.3, and illustrate the sample distributions of these variables in Figure 2.8. On the left, we see that consumption of good 1 is approximately increasing in normalized expenditure and decreasing in the normalized price, on average. On the right, we see that the sample distribution of normalized designs z_{it} is approximately log-normal.

2.6.2 Estimation of Demand

Let us now consider the estimation of the distribution of the demand random field $X_1(\cdot)$. Figure 2.9 illustrates the Nadaraya-Watson estimate $\hat{x}_1(\cdot)$ of the expected demand field $x_1(\cdot)$. In this figure, we

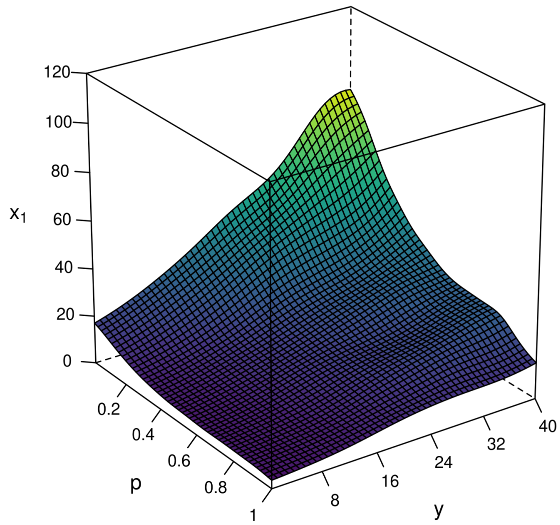

Figure 2.9. Nadaraya-Watson estimate $\hat{x}_1(\cdot)$ of expected demand $x_1(\cdot)$.

see that this estimate is strictly increasing in expenditure y and strictly decreasing in the price p, as expected when goods are normal. The shape of this surface is similar to the shape of the Stone-Geary demand function in Figure 2.2. Figure 2.10 illustrates the 95% confidence bands for this estimate. In this figure, we see that the estimate is very precise for small values of expenditure, and that the confidence bands widen as we increase expenditure. We also see that the estimate and the confidence bands are decreasing in the price, as expected. Figure 2.11 illustrates the kernel estimate of one pairwise Laplace transform of the demand random field. Of course, under a small-sigma assumption, the demand random field is approximately Gaussian, implying that it is characterized by its pairwise distributions, which are themselves characterized by their pairwise Laplace transforms. The explicit estimation of the covariance operator of the Gaussian copula is omitted for brevity. Below, I estimate the covariance operator $C(\cdot)$ of $\log M(\cdot)$ under the small-sigma assumption, and the Slutsky coefficient $\Delta_x(\cdot)$. Together, these estimates will lead to a natural estimate of the covariance operator $C_X(\cdot)$ of $X_1(\cdot)$, with the form in Corollary 2.1.

2.6.3 Estimation of Individual Preferences

Let us now consider the estimation of the distribution of individual preferences $M(\cdot)$ under the small-sigma assumption. In this section, I first estimate $\mu(\cdot)$, and then estimate the covariance operator $C(\cdot)$ of $\log M(\cdot)$ under two parametric assumptions.

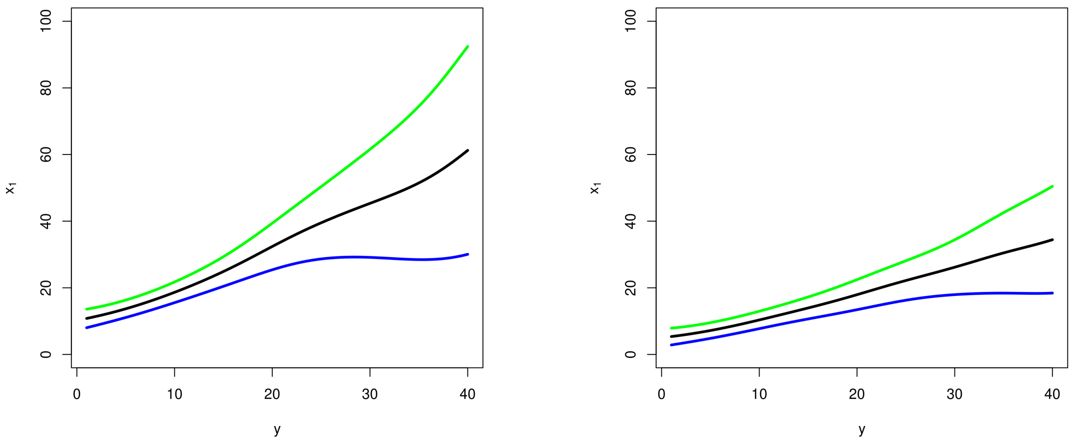

Figure 2.10. The 95% confidence bands for the Nadaraya-Watson estimate $\hat{x}_1(\cdot)$ of expected demand $x_1(\cdot)$ given $p = 0.25$ (left) and $p = 0.50$ (right). In each plot, the black line is the estimate of expected demand, the green line is the upper bound, and the blue line is the lower bound.

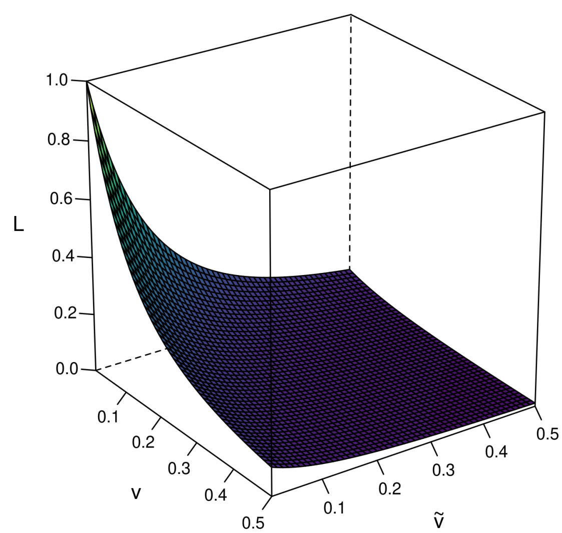

Figure 2.11. Nadaraya-Watson estimate of the pairwise Laplace transform of the demand random field given $z = (10, 0.25)$ and $\tilde{z} = (10, 0.75)$.

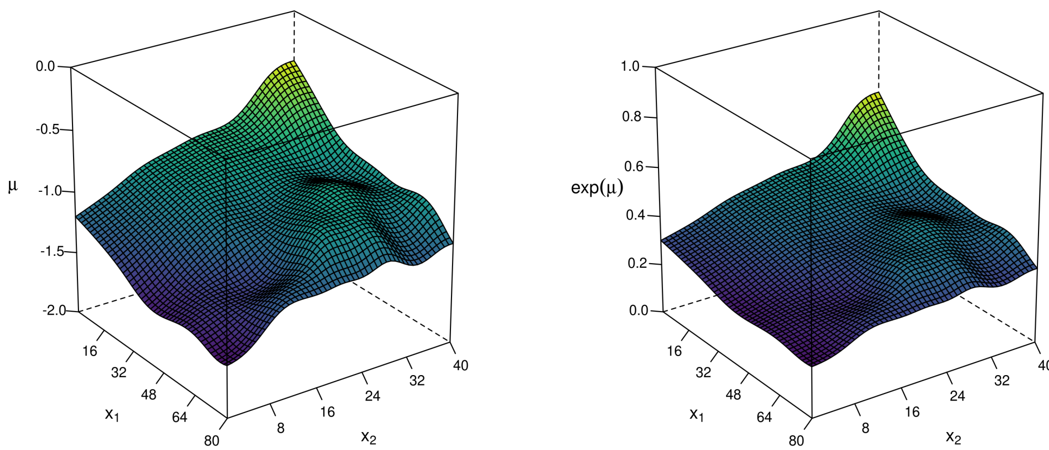

Figure 2.12. Nadaraya-Watson estimate $\hat{\mu}(\cdot)$ (left) and its exponential transform (right).

Estimation of the Mean

Under the small-sigma assumption, the functional parameter $\mu(\cdot)$ can be estimated using the Nadaraya-Watson approach, as described in Section 2.4.3. The resulting estimate $\hat{\mu}(\cdot)$ is illustrated on the left of Figure 2.12. By applying the exponential transform $\exp(\cdot)$ to this estimated field, we obtain an estimate for the expected marginal rate of substitution field $\tilde{m}(\cdot)$. The transformed estimate $\exp \hat{\mu}(\cdot)$ is illustrated on the right of Figure 2.12.

Estimation of the Covariance Operator

Next, consider the estimation of the covariance operator $C(\cdot)$ of $\log M(\cdot)$. Since we have already estimated the functional parameter $\mu(\cdot)$, we are left with the task of estimating $\theta = (\sigma_1, \sigma_2, k_1, k_2)'$, where k_j denotes the vector of parameters that characterize the covariance operator $C_j(\cdot)$ of $U_j(\cdot)$. In this section, I consider two parameterizations of $C(\cdot)$. In the first parameterization, I use Ornstein-Uhlenbeck errors:

$$C_j(x, \tilde{x}) = \exp\big\{ - k_j |x_j - \tilde{x}_j| \big\}, \tag{2.6.1}$$

and, in the second parameterization, I use squared-exponential errors:

$$C_j(x, \tilde{x}) = \exp\big\{ - k_j |x_j - \tilde{x}_j|^2 \big\}. \tag{2.6.2}$$

The results are provided in Table 2.4. In this table, we see that $\hat{\sigma}_1$ is approximately 0.38, and $\hat{\sigma}_2$ is approximately 0.44, regardless of the parameterization. These estimates are mostly consistent with the

Table 2.4. Estimated parameters $\hat{\theta}$ under Ornstein-Uhlenbeck errors and squared-exponential errors.

	Ornstein-Uhlenbeck	Squared Exponential
σ_1	0.3797	0.3875
σ_2	0.4426	0.4344
k_1	10.1283	20.9043
k_2	125.3043	27.1795

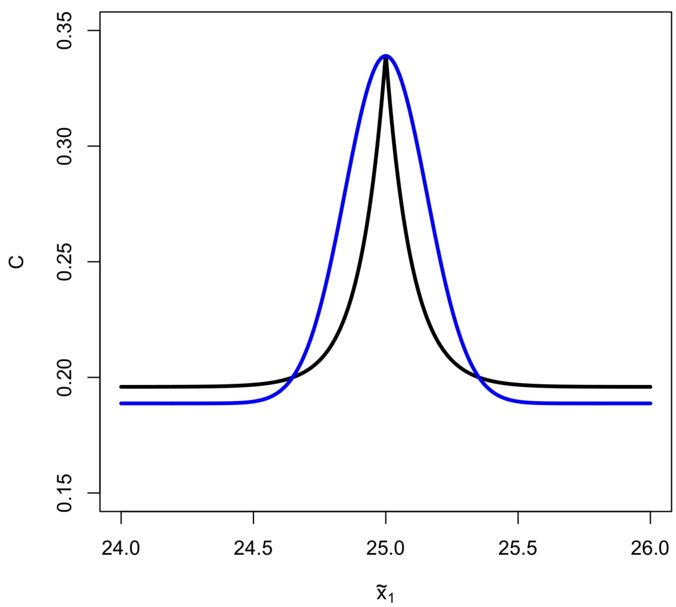

Figure 2.13. Estimated covariance operator $C(\cdot)$ of $\log M(\cdot)$ under Ornstein-Uhlenbeck errors (black) and squared-exponential errors (blue) given $x = (25, 10)$ and $\tilde{x} = (\tilde{x}_1, 10)$, with $\tilde{x}_1$ varying.

small-sigma assumption:

$$\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2} \simeq \sqrt{0.38^2 + 0.44^2} = 0.5813. \tag{2.6.3}$$

In this table, we also see that $\hat{k}_1$ is smaller than $\hat{k}_2$, regardless of the parameterization. Since, in each parameterization, k_j is a measure of dependency across quantities for good j, this result suggests that shocks to preferences for beer are less persistent across quantities than shocks to preferences for wine and liquor. Therefore, consumers are prone to liking or disliking wine and liquor at all quantities, but we cannot make a similar claim for beer. The estimated covariance operators are illustrated in Figure 2.13.

2.6.4 Estimation of the Slutsky Coefficient

The estimate of the (functional) Slutsky coefficient, based on a local quadratic fit, is illustrated on the left of Figure 2.14. For comparison, the Slutsky coefficient associated with a Stone-Geary specification is

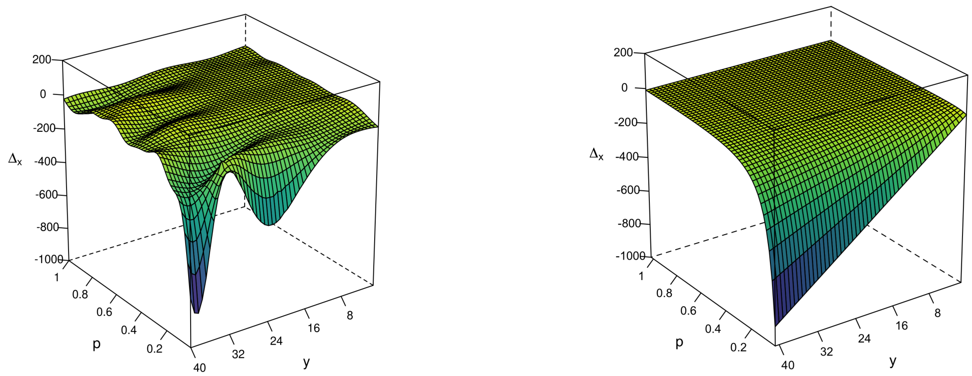

Figure 2.14. The estimated Slutsky coefficient (left) and the Stone-Geary Slutsky coefficient given $\alpha = 3/4$ (right).

given on the right. Both surfaces are relatively flat over the majority of the range of designs with a large decrease in the corner associated with large expenditures and small prices. Since the strict negativity of the Slutsky coefficient is both necessary and sufficient for integrability, this estimate can be used to evaluate the recoverability of the marginal rate of substitution $m(\cdot)$ associated with expected demand $x_1(\cdot)$ without a small-sigma assumption. Figure 2.15 illustrates the set on which the estimated Slutsky coefficient is negative.

2.6.5 Counterfactual Analysis

Now that we have estimated the expected demand field $x_1(\cdot)$, the covariance operator $C(\cdot)$ of $\log M(\cdot)$, and the Slutsky coefficient $\Delta_x(\cdot)$, we can perform individual-level counterfactual analysis under the small-sigma assumption. Figure 2.16 illustrates the counterfactual prediction with 95% confidence bands for a specific household under Ornstein-Uhlenbeck errors and squared-exponential errors. In each panel, the price has been fixed at $p = 0.0693$, and the red point denotes the observed consumption of this household in August of 2016 given $z_{it} = (7.7612, 0.0693)$. We see that, regardless of the parameterization of $C(\cdot)$, (i) the prediction passes through this point, (ii) the prediction tends to the estimated expected demand field as we move away from this point, and (iii) the width of the confidence bands tends to zero at this point. Intuitively, the conditional distribution of $X_{1i}(\cdot)$ at z_{it} is degenerate because we observe $x_{i1t} = X_{1i}(z_{it})$.

2.7 Concluding Remarks and Further Extensions

This chapter develops a non-parametric model of consumption for scanner data without the issues associated with applying traditional approaches, such as the assumptions of separability, finite-dimensional het-

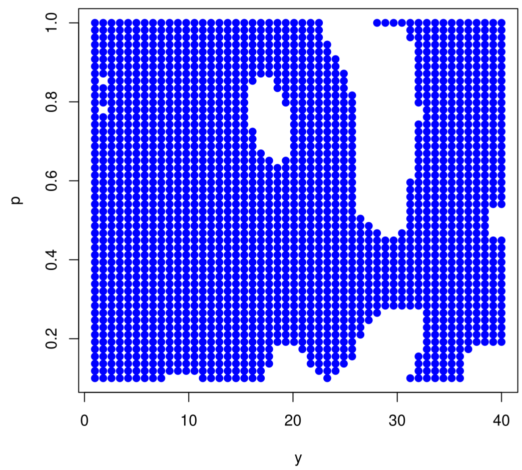

Figure 2.15. The set on which the estimated Slutsky coefficient is strictly negative.

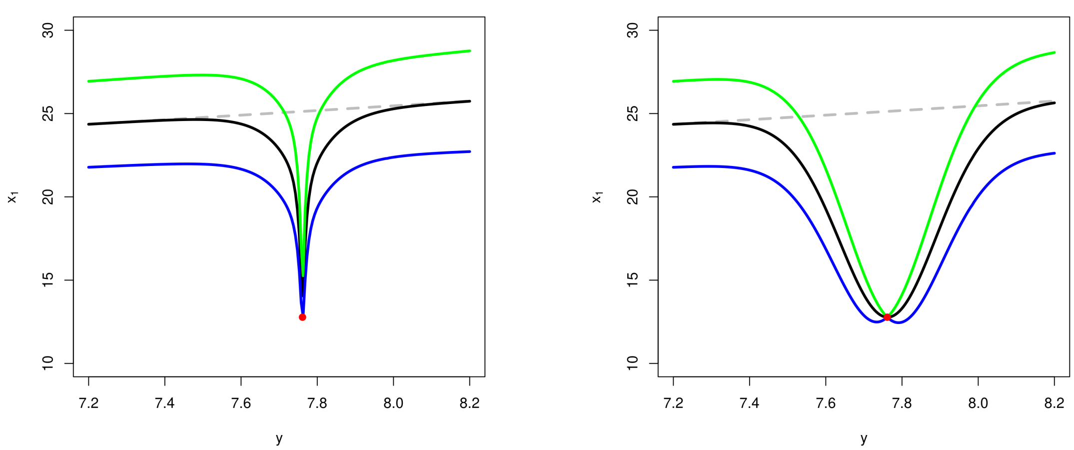

Figure 2.16. Counterfactual prediction with 95% confidence bands for a specific household given $p = 0.0693$ under Ornstein-Uhlenbeck errors (left) and squared-exponential errors (right). The dashed line is the estimated expected demand field $\hat{x}_1(\cdot)$, the black line is the mean of the distribution of $X_{1i}(\cdot)$, the blue line is the lower bound, the green line is the upper bound, and the red point is one observation for this household.

erogeneity, or monotonicity. Infinite-dimensional heterogeneity is introduced by replacing the marginal rate of substitution with a log-normal random field. This model is used to recover the latent distribution of preferences in the population, perform counterfactual analysis, test the integrability of the expected demand field at a parametric rate, and recover the preferences associated with this field. If variation in preferences is small, preferences can be recovered by approximating the relationship between demand and preferences using a first-order expansion. Else, preferences can be recovered numerically. Finally, I use the Nielsen Homescan Consumer Panel (NHCP) to illustrate these methods in an application to the consumption of alcohol.

There are several natural extensions to the model in this chapter: First, we could permit more general forms of heterogeneity by relaxing the additive separability of the error term in the construction of the marginal rate of substitution random field, or by replacing the log-normal random field with a more general process. Second, if we wanted every realization of the marginal rate of subsitution to be the marginal rate of substitution associated with some *well-behaved* utility function, without imposing any additional restrictions on heterogeneity, we could attempt to characterize the subset of processes that satisfy this restriction. Third, I do not currently permit corner solutions, intertemporal decisions, context effects, or endogenous designs, beyond what presents itself as weak dependence. Some of these features will be addressed in the subsequent chapter.

2.A Proofs

2.A.1 Proof of Lemma 2.1

By Proposition 7 in Ginsberg (1973), Theorem 5 in Arrow and Enthoven (1961), and Theorem 11.2 in Barten and Böhm (1993), the properties in (i) and (iii) are equivalent. Hence, it is sufficient to show that (ii) is equivalent to (iii) which is equivalent to (iv).

The determinant of the bordered Hessian is:

$$\Delta_u(x) = 2u_1u_2u_{12} - u_1^2u_{22} - u_2^2u_{11}, \tag{2.A.1}$$

where $u_i = u_i(x)$. Moreover:

$$\frac{d^2g(x_1,u)}{dx_1^2} = \Delta_m(x) = \frac{1}{u_2^3}\Delta_u(x), \tag{2.A.2}$$

where $x_2 = g(x_1, v)$. The result follows from the strict positivity of u_2 and the fact that, for every $x \in R$, there exists some $v \in \mathbb{R}$ that satisfies $x_2 = g(x_1, v)$.

2.A.2 Single-Valued and Continuously-Differentiable Demand

If a quantity $x_1 > 0$ satisfies (2.2.9) given $z \in R$, then:

$$\frac{dm}{dx_1}(x_1, y - px_1) = \frac{\partial m}{\partial x_1}(x) - m(x)\frac{\partial m}{\partial x_2}(x) = -\Delta_m(x). \tag{2.A.3}$$

This derivative is strictly negative by Lemma 2.1. Consequently, the implicit function theorem implies that there exists an open set U containing $z_0 \in R$, and a unique and continuously-differentiable implicit function $x_1^*(\cdot)$ on U that satisfies:

$$x_1^*(z_0) = x_1 \text{ and } m(x_1^*(z), y - px_1^*(z)) - p = 0, \tag{2.A.4}$$

for every $z \in U$. This solution can be extended to R, if $x_1^*(z)$ is the only quantity that satisfies (2.2.9) given z. Suppose that there exists another quantity $x_1^* < x_1^*(z)$ that satisfies (2.2.9) given z, and define:

$$\gamma(t) = m\big[tx_1^* + (1-t)x_1^*(z), y - p(tx_1 + (1-t)x_1^*(z))\big] - p, \tag{2.A.5}$$

for $t \in [0,1]$.[17] This function satisfies $\gamma(t_1) < 0 < \gamma(t_0)$, for some $t_0 < t_1$, since:

(i) x_1 and $x_1^*(z)$ satisfy (2.2.9) given z

(ii) $\gamma(t)$ is continuous

(iii) the derivative in (2.A.3) is strictly negative at every $x_1 > 0$ that satisfies (2.2.9).

Therefore, the intermediate value theorem implies that there exists $t^* \in (t_0, t_1)$ such that $\gamma(t^*) = 0$ and $\gamma'(t^*) \leq 0$. However, $\gamma'(t^*) \leq 0$ implies that there exists $x_1 \neq x_1(z)$ that satisfies (2.2.9), but fails to satisfy (2.A.3), generating a contradiction. See Figure 2.17 for an illustration. This use of the intermediate value theorem is specific to the two good framework. Otherwise, more complex arguments are required (see, for example, Gale and Nikaidô, 1965, and Mas-Colell, 1979).

2.A.3 Slutsky Matrix Properties

The $(i,j)^{th}$-entry of the 2-by-2 Slutsky matrix is defined by:

$$S_{ij}(\tilde{z}) \equiv \frac{\partial \tilde{x}_i}{\partial \tilde{p}_j}(\tilde{z}) + \tilde{x}_j(\tilde{z})\frac{\partial \tilde{x}_i}{\partial \tilde{y}}(\tilde{z}), \tag{2.A.6}$$

in which $\tilde{z} = (\tilde{p}_1, \tilde{p}_2, \tilde{y}) \in \mathbb{R}^3_{++}$, and $\tilde{x}(\cdot)$ denotes the demand field as a function of $\tilde{z}$, prior to any normalizations. I show that, if $\tilde{x}(\cdot)$ is homogeneous of degree zero, and it satisfies Walras' law, then, in the two good framework: (i) the Slutsky matrix is symmetric, (ii) $\Delta_x(z) \leq 0$ if, and only if, the Slutsky

[17] The assumption that $x_1^* < x_1^*(z)$ is without loss of generality.

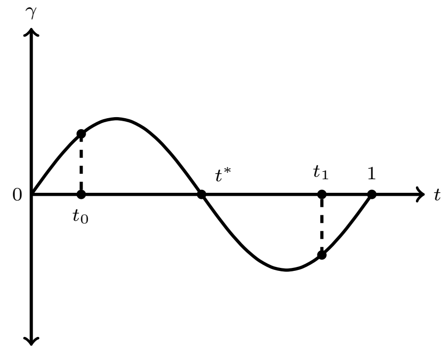

Figure 2.17. The intuition in the proof of continuous-differentiability.

matrix is negative semi-definite, and (iii) $\Delta_x(z) < 0$ implies that the Slutsky matrix has exactly one strictly negative eigenvalue.

(i) **Symmetry:** Symmetry holds if, and only if:

$$\frac{\partial \tilde{x}_1}{\partial \tilde{p}_2}(\tilde{z}) + \tilde{x}_2(\tilde{z})\frac{\partial \tilde{x}_1}{\partial \tilde{y}}(\tilde{z}) = \frac{\partial \tilde{x}_2}{\partial \tilde{p}_1}(\tilde{z}) + \tilde{x}_1(\tilde{z})\frac{\partial \tilde{x}_2}{\partial \tilde{y}}(\tilde{z}). \tag{2.A.7}$$

By Walras' law, we have:

$$\tilde{x}_2(\tilde{z}) = \frac{\tilde{y} - \tilde{p}_1\tilde{x}_1(\tilde{z})}{\tilde{p}_2}. \tag{2.A.8}$$

Now, by replacing $\tilde{x}_2(\tilde{z})$ in (2.A.7) and rearranging, we get:

$$\frac{\partial \tilde{x}_1}{\partial \tilde{p}_2}(\tilde{z}) + \frac{\tilde{y}}{\tilde{p}_2}\frac{\partial \tilde{x}_1}{\partial \tilde{y}}(\tilde{z}) = -\frac{\tilde{p}_1}{\tilde{p}_2}\frac{\partial \tilde{x}_1}{\partial \tilde{p}_1}(\tilde{z}). \tag{2.A.9}$$

As a consequence, the Slutsky martix is symmetric if, and only if, this equality holds, or equivalently, if, and only if:

$$\tilde{p}_1\frac{\partial \tilde{x}_1}{\partial \tilde{p}_1}(\tilde{z}) + \tilde{p}_2\frac{\partial \tilde{x}_1}{\partial \tilde{p}_2}(\tilde{z}) + \tilde{y}\frac{\partial \tilde{x}_1}{\partial \tilde{y}}(\tilde{z}) = 0. \tag{2.A.10}$$

This equality is satisfied by Euler's Theorem for homogeneous functions.

(ii) **Negative Semi-Definiteness:** The homogeneity of demand $\tilde{x}(\cdot)$ implies that the Slutsky matrix is singular. Therefore, the rank of the Slutsky matrix is strictly smaller than two, and it has at least one eigenvalue that is equal to zero. Since the Slutsky matrix is symmetric by (i), the other eigenvalue is equal to the trace of the Slutsky matrix:

$$\frac{\partial \tilde{x}_1}{\partial \tilde{p}_1}(\tilde{z}) + \tilde{x}_1(\tilde{z})\frac{\partial \tilde{x}_1}{\partial \tilde{y}}(\tilde{z}) + \frac{\partial \tilde{x}_2}{\partial \tilde{p}_2}(\tilde{z}) + \tilde{x}_2(\tilde{z})\frac{\partial \tilde{x}_2}{\partial \tilde{y}}(\tilde{z}). \tag{2.A.11}$$

By applying (2.A.8), this trace is weakly less than zero if, and only if:

$$\left(1 + p^2\right)\left[\frac{\partial x_1}{\partial p}(z) + x_1(z)\frac{\partial x_1}{\partial y}\right] \leq 0. \tag{2.A.12}$$

Equivalently, if, and only if, $\Delta_x(z) \leq 0$. Because a symmetric matrix is negative semi-definite if, and only if, all of its eigenvalues are non-positive, the Slutsky matrix is negative semi-definite if, and only if, $\Delta_x(z) \leq 0$.

(iii) **Negative Eigenvalue:** By the argument above, if $\Delta_x(z) < 0$, the trace of the Slutsky matrix is strictly negative, and one eigenvalue is strictly negative.

2.A.4 Proof of Proposition 2.1

Properties (i) and (ii) follow from Lemma 2.2. The proof is similar to the proof in Appendix 2.A.2. Let us now consider property (iii). The Jacobian of the demand field $x(\cdot)$ is:

$$J_x(z) = \begin{bmatrix} \frac{\partial x_1}{\partial y}(z) & \frac{\partial x_1}{\partial p}(z) \\ \frac{\partial x_2}{\partial y}(z) & \frac{\partial x_2}{\partial p}(z) \end{bmatrix}. \tag{2.A.13}$$

By Walras' law:

$$J_x(z) = \begin{bmatrix} \frac{\partial x_1}{\partial y}(z) & \frac{\partial x_1}{\partial p}(z) \\ 1 - p\frac{\partial x_1}{\partial y}(z) & -x_1(z) - p\frac{\partial x_1}{\partial p}(z) \end{bmatrix}. \tag{2.A.14}$$

The determinant of this matrix is $-\Delta_x(z)$. Similarly, the Jacobian of the inverse demand function $z(\cdot)$ is defined by:

$$J_z(x) = \begin{bmatrix} \frac{\partial y}{\partial x_1}(x) & \frac{\partial y}{\partial x_2}(x) \\ \frac{\partial p}{\partial x_1}(x) & \frac{\partial p}{\partial x_2}(x) \end{bmatrix}. \tag{2.A.15}$$

Again, by Walras' law:

$$J_z(x) = \begin{bmatrix} x_1\frac{\partial p}{\partial x_1}(x) + p(x) & x_1\frac{\partial p}{\partial x_2}(x) + 1 \\ \frac{\partial p}{\partial x_1}(x) & \frac{\partial p}{\partial x_2}(x) \end{bmatrix}. \tag{2.A.16}$$

Since, under Assumption 2.1, we have: $m(x) = p(x)$, for every $x \in R$, the determinant of this matrix is $\Delta_m(x)$, and the determinant of its inverse is $\Delta_m(x)^{-1}$. The remainder of the proof follows from the inverse function theorem.

2.A.5 Proof of Theorem 2.1

Part (i) follows directly from Proposition 2.1(i). Part (ii) follows from (a) the Picard-Lindelöf theorem, (b) the fact that, under Assumption A, the solution to the differential equation in (2.2.2) subject to $g(x_1^*, u) = x_2^*$ can be analytically extended to the boundary of $\mathcal{X}$, and (c) the fact that, by the continuity of the indifference curves, we can extend our knowledge of each indifference curve over some open set to its closure.

2.A.6 Proof of Proposition 2.2

(i) The random field $M(\cdot)$ is log-normal because $\log M(\cdot)$ is Gaussian.

(ii) By applying the Laplace transform of a Gaussian distribution, we obtain:

$$\tilde{m}(x) = \mathbb{E}[\exp\{\log M(x)\}] = \exp\big\{\mathbb{E}[\log M(x)] + \frac{1}{2}V[\log M(x)]\big\}. \tag{2.A.17}$$

(iii) Define: $\gamma(x, \tilde{x}) = \exp\big\{\mu(x) + \mu(\tilde{x})\big\}$. By applying the Laplace transform again:

$$\begin{aligned}\mathbb{E}[M(x)M(\tilde{x})] &= \gamma(x,\tilde{x})\mathbb{E}\big[\exp\big\{\sigma_1 U_1(x_1) + \sigma_2 U_2(x_2) + \sigma_1 U_1(\tilde{x}_1) + \sigma_2 U_2(\tilde{x}_2)\big\}\big] \\ &= \gamma(x,\tilde{x})\exp\big\{\frac{1}{2}V\big[\sigma_1 U_1(x_1) + \sigma_2 U_2(x_2) + \sigma_1 U_1(\tilde{x}_1) + \sigma_2 U_2(\tilde{x}_2)\big]\big\} \\ &= \gamma(x,\tilde{x})\exp\big\{\sigma_1^2 + \sigma_2^2\big\}\exp\big\{C(x,\tilde{x})\big\}.\end{aligned} \tag{2.A.18}$$

Similarly, we can write:

$$\mathbb{E}[M(x)]\mathbb{E}[M(\tilde{x})] = \gamma(x,\tilde{x})\exp\big\{\sigma_1^2 + \sigma_2^2\big\}. \tag{2.A.19}$$

Together, these results yield the form of $C_M(\cdot)$ in (2.3.5).

2.A.7 Proof of Proposition 2.3

(i) By performing a first-order expansion, we get:

$$\begin{aligned}M(x) &= \exp\big\{\log M(x)\big\} \\ &= \exp\big\{\mu(x) + \sigma_1 U_1(x_1) + \sigma_2 U_2(x_2)\big\} \\ &= \exp\big\{\mu(x)\big\}\exp\big\{\sigma_1 U_1(x_1) + \sigma_2 U_2(x_2)\big\} \\ &= \exp\big\{\mu(x)\big\}\big[1 + \sigma_1 U_1(x_1) + \sigma_2 U_2(x_2)\big] + o(\sigma) \\ &= m(x) + m(x)\big[\sigma_1 U_1(x_1) + \sigma_2 U_2(x_2)\big] + o(\sigma).\end{aligned} \tag{2.A.20}$$

Since $U_1(\cdot)$ and $U_2(\cdot)$ are Gaussian processes, the random field $M(\cdot)$ is approximately Gaussian.

(ii) The effect on $M(\cdot)$ of a small perturbation of $X_1(\cdot)$ is:

$$\delta M\big[x(z)\big] = \left[\frac{\partial m}{\partial x_1}(x(z)) - p\frac{\partial m}{\partial x_2}(x(z))\right]\delta X_1(z) = \Delta_m\big[x(z)\big]\delta X_1(z). \tag{2.A.21}$$

Therefore, we obtain:

$$\delta X_1(z) = \Delta_m\big[x(z)\big]^{-1}\delta M\big[x(z)\big] = -\Delta_x(z)\delta M\big[x(z)\big], \tag{2.A.22}$$

and we can deduce the following small-sigma approximation of $X_1(\cdot)$:

$$\begin{aligned} X_1(z) &= x_1(z) + \delta X_1(z) \\ &= x_1(z) - \Delta_x(z)\delta M(x(z)) \\ &= x_1(z) - \Delta_x(z)\big\{m(x(z))\big[\sigma_1 U_1(x_1(z)) + \sigma_2 U_2(x_2(z))\big] + o(\sigma)\big\} \\ &= x_1(z) - p\Delta_x(z)\big[\sigma_1 U_1(x_1(z)) + \sigma_2 U_2(x_2(z))\big] + o(\sigma). \end{aligned} \tag{2.A.23}$$

Consequently, the demand field is approximately Gaussian.

(iii) If $M(\cdot)$ is continuously-differentiable, then $G_0(\cdot, v)$ is well-defined by the Picard-Lindelöf theorem. If $\sigma_1 = \sigma_2 = 0$, then $G_0(x_1, v) = g_0(x_1, v)$. We are, therefore, looking for an expansion of $G_0(\cdot, v)$ with the form:

$$G_0(x_1, v) = g_0(x_1, v) + h_0(x_1, v; \sigma, U_1, U_2) + o(\sigma), \tag{2.A.24}$$

where $h_0(x_{10}, u; \sigma, U_1, U_2) = 0$. By differentiating with respect to x_1, we obtain:

$$\begin{aligned} \frac{\partial G_0(x_1, v)}{\partial x_1} &= -M(x_1, G_0(x_1, v)) \\ = -m(x_1, g_0(x_1, v)) &- \frac{\partial m}{\partial x_2}\big[x_1, g_0(x_1, v)\big] h_0(x_1, v; \sigma, U_1, U_2) \\ &+ m(x_1, g_0(x_1, v))\big[\sigma_1 U_1(x_1) + \sigma_2 U_2(g_0(x_1, v))\big] + o(\sigma). \end{aligned} \tag{2.A.25}$$

By (2.A.24) and the definition of $g_0(x_1, v)$, we deduce:

$$\begin{aligned} \frac{\partial h_0(x_1, v; \sigma, U_1, U_2)}{\partial x_1} = &- \frac{\partial m}{\partial x_2}\big[x_1, g_0(x_1, v)\big] h_0(x_1, v; \sigma, U_1, U_2) \\ &+ m(x_1, g_0(x_1, v))\big[\sigma_1 U_1(x_1) + \sigma_2 U_2(g_0(x_1, v))\big]. \end{aligned} \tag{2.A.26}$$

With the initial condition $h_0(x_{10}, v; \sigma, U_1, U_2) = 0$, this differential equation has a unique solution. We are, therefore, left with the form in part (iii).

2.A.8 Proof of Corollary 2.2

By Proposition 2.4, applying the delta-method yields:

$$\sqrt{nTh_1h_2}\big[\hat{x}(z) - x(z)\big] \xrightarrow{d} N\left(0, \kappa\frac{\sigma^2(z)}{\pi(z)}\begin{pmatrix} 1 & -p \\ -p & p^2 \end{pmatrix}\right). \tag{2.A.27}$$

Since $\Delta_x(z) \neq 0$ coincides with the determinant of the Jacobian of expected demand, the inverse function theorem implies that expected demand is invertible in a neighbourhood of z. Because $z(x(z)) = z$ and

$m(x) = p(x)$, applying the delta-method a second time yields:

$$\sqrt{nTh_1h_2}\big[\hat{z}(x) - z(x)\big] \xrightarrow{d} N\left(0, \kappa\frac{\sigma_z^2(x)}{\pi_z(x)}J_z(x)\begin{pmatrix} 1 & -m(x) \\ -m(x) & m(x)^2 \end{pmatrix}J_z(x)'\right). \tag{2.A.28}$$

where $J_z(x)$ denotes the Jacobian of $z(\cdot)$. Thus, the second component of $z(\cdot)$ satisfies:

$$\sqrt{nTh_1h_2}\,[\hat{m}(x) - m(x)] \xrightarrow{d} N\left(0, \kappa\frac{\sigma_z^2(x)}{\pi_z(x)}\left[\frac{\partial m(x)}{\partial x_1} - m(x)\frac{\partial m(x)}{\partial x_2}\right]^2\right) \tag{2.A.29}$$

The result follows because the expression in the parentheses is equal to the negative of $\Delta_m(x)$, and Proposition 2.1(iii) implies $\Delta_m(x) = -\Delta_x(z(x))^{-1}$.

2.B Summary Statistics

In this appendix, I summarize some demographics of the households in the Nielsen Homescan Consumer Panel (NHCP), after restricting our sample to households with strictly positive expenditure in each of the two aggregate groups of alcoholic beverages in each month from August to November in 2016. While the NHCP sample is balanced to reflect the population of households in the United States by household size, income, age, education, and race, the selected sample may not be balanced. I compare the demographics in the selected NHCP sample to the Current Population Survey (CPS). I refer the reader to Guha and Ng (2019) for additional summary statistics.

Table 2.5 describes the distribution of household size in the selected sample and the CPS. The majority of households in the selected NHCP sample have two members. The shape of the size distribution in the selected sample is similar to the shape of the size distribution in the CPS, with a smaller proportion of one-member households, and a larger proportion of two-member households. This difference could be explained by (i) one-member households having difficulty meeting the required level of spending to be included in the selected sample, or (ii) two-member households being more likely to consume alcohol in both aggregate groups.

Table 2.5. Household size in the selected NHCP sample and the 2017 Annual Social and Economic Supplement (ASEC) of the CPS. CPS numbers are in thousands.

	Sample		CPS	
Size	Number	Proportion	Number	Proportion
1	122	0.1578	35,388	0.2812
2	442	0.5717	42,785	0.3400
3	112	0.1448	19,423	0.1543
4	69	0.0892	16,267	0.1292
5	18	0.0232	7,548	0.0599
6	6	0.0077	2,813	0.0223
7+	4	0.0051	1,596	0.0126
Total	773	1.0000	125,819	1.0000

Table 2.6 describes the distribution of household size in the selected NHCP sample and the CPS. These two distributions are similar. The only noticable difference is at the upper tail: The selected sample includes a higher proportion of households earning \$70,000 to \$99,999, and a smaller proportion of households earning \$100,000 or more.

Table 2.6. Annual household income in the selected NHCP sample and the 2017 Annual Social and Economic Supplement (ASEC) of the CPS. CPS numbers are in thousands.

	Sample		CPS	
Income	Number	Proportion	Number	Proportion
Under $5,000	10	0.0129	4,138	0.0327
$5,000 to $9,999	9	0.0116	3,878	0.0307
$10,000 to $14,999	17	0.0219	6,122	0.0485
$15,000 to $19,999	25	0.0323	5,838	0.0462
$20,000 to $24,999	35	0.0452	6,245	0.0494
$25,000 to $29,999	41	0.0530	5,939	0.0470
$30,000 to $34,999	45	0.0582	5,919	0.0468
$35,000 to $39,999	38	0.0491	5,727	0.0453
$40,000 to $44,999	32	0.0413	5,487	0.0434
$45,000 to $49,999	35	0.0452	5,089	0.0403
$50,000 to $59,999	92	0.1190	9,417	0.0746
$60,000 to $69,999	55	0.0711	8,213	0.0650
$70,000 to $99,999	169	0.2186	19,249	0.1524
$100,000+	170	0.2199	34,963	0.2769
Total	773	1.0000	126,224	1.0000

Tables 2.7 and 2.8 describe the distribution of the age of the head of the household in the selected NHCP sample and the householder in the CPS. I provide two tables because there is no direct comparison between the samples. Indeed, the NHCP reports the age of the household head, separated by gender, while the CPS reports the age of the householder, the person that owns or leases the residence. Table 2.7 compares the distribution of the age of the eldest household head in the selected sample with the age of the householder in the CPS. The age of the eldest head in the selected sample is much more concentrated around the ages of 50 to 74. This result might be driven by the fact that, by definition, the eldest head is always (at least weakly) older than the householder, but this result persists in Table 2.8, in which the selected sample is separated by gender.

Table 2.7. Age of eldest household head in the selected NHCP sample and the householder in the 2017 Annual Social and Economic Supplement (ASEC) of the CPS. CPS numbers are in thousands.

	Sample		CPS	
Age	Number	Proportion	Number	Proportion
Under 20	0	0.0000	753	0.0059
20 to 24	2	0.0025	5,608	0.0445
25 to 29	4	0.0051	9,453	0.0751
30 to 34	24	0.0310	10,594	0.0842
35 to 39	33	0.0426	10,651	0.0846
40 to 44	35	0.0452	10,571	0.0840
45 to 49	44	0.0569	11,115	0.0883
50 to 54	81	0.1047	12,180	0.0968
55 to 64	271	0.3505	23,896	0.1899
65 to 74	222	0.2871	17,551	0.1394
75+	57	0.0737	13,448	0.1068
Total	773	1.0000	125,819	1.0000

Table 2.8. Age of household head by gender in the selected NHCP sample and the householder in the 2017 Annual Social and Economic Supplement (ASEC) of the CPS. CPS numbers are in thousands.

	Male		Female		CPS	
Age	Number	Proportion	Number	Proportion	Number	Proportion
Under 20	0	0.0000	0	0.0000	753	0.0059
20 to 24	1	0.0014	2	0.0029	5,608	0.0445
25 to 29	6	0.0089	8	0.0118	9,453	0.0751
30 to 34	21	0.0313	30	0.0443	10,594	0.0842
35 to 39	29	0.0433	39	0.0576	10,651	0.0846
40 to 44	35	0.0523	29	0.0428	10,571	0.0840
45 to 49	42	0.0627	49	0.0724	11,115	0.0883
50 to 54	75	0.1121	97	0.1434	12,180	0.0968
55 to 64	240	0.3587	233	0.3446	23,896	0.1899
65 to 74	172	0.2571	159	0.2352	17,551	0.1394
75+	48	0.0717	30	0.0443	13,448	0.1068
Total	669	1.0000	676	1.0000	125,819	1.0000

Table 2.9 describes the distribution of the marital status of the head of the household in the selected NHCP sample and the marital status of people aged 15 and over in the CPS. We observe a higher proportion of married households and a lower proportion of single households. This result is consistent with the fact that we observe a smaller proportion of one-member households, and a larger proportion of two-member households. These samples have approximately the same proportions of widowed and divorced/separated households.

Table 2.9. Marital status of the head of the household in the selected NHCP sample and the marital status of people aged 15 and over in the 2017 Annual Social and Economic Supplement (ASEC) of the CPS. CPS numbers are in thousands.

	Sample		CPS	
Marital Status	Number	Proportion	Number	Proportion
Married	588	0.7606	130,606	0.5041
Widowed	34	0.0439	14,919	0.0575
Divorced/Separated	79	0.1021	30,626	0.1182
Single	72	0.0931	82,912	0.3200
Total	773	1.0000	259,063	1.0000

Table 2.10 describes the distribution of the (self-reported) racial identity of household in the selected NHCP sample and the race of the householder in the CPS. Table 2.10 uses the categories in the NHCP. The CPS categories are (i) White Alone, (ii) White Alone (Non-Hispanic), (iii) Black Alone, (iv) Asian Alone, and (v) Hispanic (Any Race). Unclassified observations in the CPS are in a sixth category: Other. I reclassify category (ii) as White/Caucasian, category (iii) as Black/African American, and category (iv) as Asian, then classify the remaining households as Other. I observe a higher proportion of White/Caucasian households.

Table 2.10. Racial identity of household in the selected NHCP sample and the race of the householder in the 2017 Annual Social and Economic Supplement (ASEC) of the CPS. CPS numbers are in thousands.

	Sample		CPS	
Race	Number	Proportion	Number	Proportion
White/Caucasian	640	0.8279	84,387	0.6685
Black/African American	77	0.0996	16,733	0.1325
Asian	14	0.0181	6,392	0.0506
Other	42	0.0543	18,712	0.1482
Total	773	1.0000	126,224	1.0000

Table 2.11 describes the employment status of the head of the household in the selected NHCP sample by gender and the employment status of people in the CPS. The No Head category characterizes households in the selected sample that do not have a male or female head, as reported by the panelist. I reclassify households in the CPS that are employed, but not "at work," as Unemployed. We observe a lower proportion of households working at least 35 hours a week and a higher proportion of unemployed households.

Table 2.11. Employment status of the head of the household in the selected NHCP sample by gender and employment status of people aged 16 and over in the 2017 Labor Force Statistics from the CPS. CPS numbers are in thousands.

	Male		Female		CPS	
Weekly Hours	Number	Proportion	Number	Proportion	Number	Proportion
Under 30	52	0.0672	87	0.1125	24,163	0.1507
30 to 34	16	0.0206	35	0.0452	10,916	0.0680
35+	346	0.4476	238	0.3078	112,651	0.7026
Unemployed	255	0.3298	316	0.4087	12,590	0.0785
No Head	104	0.1345	97	0.1254	-	-
Total	773	1.0000	773	1.0000	160,320	1.0000

Table 2.12 describes the educational attainment of the head of the household in the selected NHCP sample by gender and the educational attainment of people in the CPS. Once again, the No Head category characterizes households in the selected sample that do not have a male or female head. We observe a higher proportion of households with some college and a lower proportion of households with at least a college diploma.

Table 2.12. Educational attainment of the head of the household in the selected NHCP sample by gender and educational attainment of people aged 18 and over in the 2017 Annual Social and Economic Supplement (ASEC) of the CPS. CPS numbers are in thousands.

	Male		Female		CPS	
Education	Number	Proportion	Number	Proportion	Number	Proportion
Grade School	8	0.0103	1	0.0012	9,187	0.0372
Some HS	32	0.0413	10	0.0129	17,996	0.0730
HS Grad.	178	0.2302	198	0.2561	71,170	0.2889
Some Coll.	192	0.2483	220	0.2846	46,445	0.1885
Coll. Grad.	186	0.2406	166	0.2147	73,439	0.2981
Post-Coll. Grad.	73	0.0944	81	0.1047	7,292	0.0296
No Head	104	0.1345	97	0.1254	-	-
Total	773	1.0000	773	1.0000	246,325	1.0000

HS: High school; Coll: College.

2.C Local Quadratic Fit

In this appendix, I provide the technical conditions that are needed to estimate the partial derivatives of expected demand $x_1(\cdot)$. The following discussion follows from Theorem 4 in Section 4 in Lu (1996). As in the main text, let $\hat{a}(z)$, $\hat{b}(z)$, and $\hat{c}(z)$ denote the arguments that minimize the objective function in (2.5.11). Let $\sigma_0^2(\cdot)$ denote the conditional variance of $X_1(Z)$ given Z. Then, we obtain:

$$X_1(Z) = x_1(Z) + \sigma_0(Z)e, \tag{2.C.1}$$

in which $\mathbb{E}(e|Z) = 0$ and $V(e|Z) = 1$. This framework is the "random design" model on page 188 of Lu (1996).

The main regularity conditions are the following:

Assumption C.

(i) There exists $\delta > 0$ such that $\mathbb{E}|X_1(Z)|^{2+\delta} < \infty$.

(ii) The kernel function $K(\cdot)$ is a spherically symmetric density function.

(iii) The kernel function $K(\cdot)$ has its 12^{th}-power marginal moment:

$$\int u_j^{12} K(u_1, u_2) du_1 du_2 < \infty, \tag{2.C.2}$$

for each $j = 1, 2$.

(iv) The fourth derivative of expected demand $x_1(\cdot)$ is continuous on R.

(v) The density $\pi(\cdot)$ is continuously-differentiable on R.

(vi) The conditional variance $\sigma_0^2(\cdot)$ is continuous on R.

Assumption C is the additional assumption that we need to estimate the partial derivatives of expected demand $x_1(\cdot)$ using the local quadratic fit described in Section 2.5.3 (see Theorem 4 in Lu, 1996). The explicit expressions of $B(h, z)$ and $\Sigma(z)$ are omitted for brevity, but can be found in Theorem 3 of Lu (1996).

2.D Simulating Individual Random Fields

In Section 2.3.2, I deduced the approximate expressions for the random fields $M(\cdot)$ and $X_1(\cdot)$ when errors are small, but these results are approximate and say nothing about what happens when errors are large. In this section, I explain how to derive the random fields $M(\cdot)$ and $X_1(\cdot)$ using *simulation-based* *methods* when the underlying model (defined by the functional parameters, $\mu(\cdot)$ and $C(\cdot)$, the scalar parameter σ, and the way in which measureable solutions to (2.3.16) are chosen) is known.

Since the relationship in (2.3.16), which describes how to transform $M(\cdot)$ into $X_1(\cdot)$, does not depend on x, if we know the latent model, then realizations of $M(\cdot)$ and $X_1(\cdot)$ can be deduced by simulating a realization of $\log M(\cdot)$, applying the exponential transform to obtain a realization of $M(\cdot)$, and solving (2.3.16) to derive $X_1(\cdot)$.

Gaussian processes are easy to simulate. Of course, it is not possible to simulate the continuous trajectory itself, but we can simulate a space-discretized version. In particular, we can simulate $U_j(\cdot)$ by constructing a grid $(x_j^1, \ldots, x_j^L)$ and simulating a single draw from a zero-mean, multivariate Gaussian distribution of dimension L with a covariance matrix Σ_j whose k^{th} row and ℓ^{th} column entry equals $C_j(x_j^k, x_j^\ell)$.

Remark 2.7. In such a framework of non-linear random fields, it is important to have information about computation time. It takes approximately 36 seconds to simulate 5000 marginal rates of substitution on a 100-by-100 grid under Ornstein-Uhlenbeck errors given $k_1 = 10$, $k_2 = 2$, and $\sigma_1 = \sigma_2 = 1/2$.

2.E The Probability Space and Differentiability

In this appendix, I briefly review the Skorokhod space, a metric that can be introduced in order to make the Skorokhod space a Polish space, and the probability space in Section 2.3.2. I, then, consider the aforementioned extensions of standard results.

2.E.1 The Definition of the Skorokhod Space

The Skorokhod space $D[0,c)^2$ contains all deterministic fields on $[0,c)^2$ that are cadlag (right-continuous with left-limits). Therefore, it contains all continuous fields, some fields with jumps, and the joint cumulative distribution function of any pair of non-negative random variables. See Figure 2.18.

2.E.2 A Metric

The Skorokhod space $D[0,c)^2$ can be made into a metric space in a number of ways, but some metrics are not appropriate in practice. Typically, we are interested in approximating a field—for example, when we discretize its domain for simulations or estimate from random observations. Such approximations are manageable when the topological space is complete and separable.[18] By definition, a Polish space satisfies these properties. It is, therefore, sufficient to equip $D[0,c)^2$ with a metric that turns it into a Polish space. While it is tempting to use the supremum norm, this choice would turn $D[0,c)^2$ into a non-separable Banach space, making it far from practical, motivating the use of a different metric. I use a

[18] A topological space is complete if every Cauchy sequence converges and separable if it contains a countable dense subset.

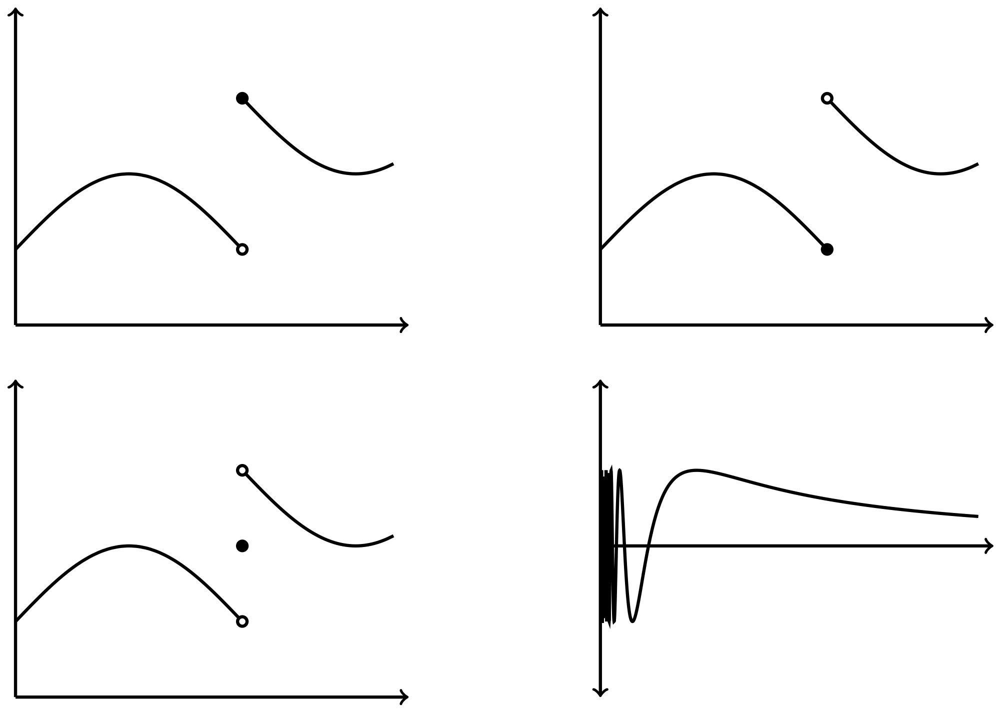

Figure 2.18. Top-left: cadlag (right-continuous with left-limits) but not caglad (left-continuous with right-limits). Top-right: caglad but not cadlad. Bottom-left: Not cadlag or caglad. Bottom-right: Plot of $\sin\left(\frac{1}{x}\right)$—a more interesting example of a function that is caglad but not cadlad.

metric that (i) coincides with the supremum norm on the subspace of continuous functions with compact support, and (ii) extends this norm in an appropriate way to all of the other elements of $D[0,c)^2$.

This metric is constructed in the following way:

(i) For continuous $f, h \in D[0,c)^2$ with compact support $\mathcal{K}$, the standard metric is:

$$d_{\mathcal{K}}(f,h) = ||f-h||_{\mathcal{K}} = \sup_{x\in\mathcal{K}} ||f(x)-h(x)||, \tag{2.E.1}$$

where $||\cdot||$ is the standard Euclidean norm.

(ii) This metric can be extended to all continuous $f, h \in D[0,c)^2$ by considering a sequence of closed balls (B_n), centred at 0, with radius n, and their intersection with the positive orthant. For continuous $f, h \in D[0,c)^2$, this procedure yields:

$$d_C(f,h) = \sum_{n=1}^{\infty} \frac{1}{2^{n-1}} \left(\frac{||f-h||_{B_n}}{1+||f-h||_{B_n}} \right). \tag{2.E.2}$$

(iii) Extending this metric to all functions in $D[0,c)^2$ requires some additional care. To illustrate, Skorokhod (1956) provided an extension of the supremum norm to $D[0,1]$, but an extension to $D[0,\infty)$ was not provided until somewhat recently in Lindvall (1973). Is make use of one of the metrics proposed by Straf (1972). For each $\varepsilon > 0$, let Λ_ε denote the set of homeomorphisms $\lambda = (\lambda_1, \lambda_2)$ that map from $[0,c)^2$ to $[0,c)^2$ where $\lambda(0) = 0$, and for each component $i = 1, 2$, we have:

$$\sup_{z_i \neq z_i'} \left| \log \frac{\lambda_i(z_i) - \lambda_i(z_i')}{z_i - z_i'} \right| < \varepsilon. \tag{2.E.3}$$

In other words, Λ_ε is the set of homeomorphisms, fixed at zero, whose components have a diffeomorphism norm that is bounded above by ε. Now, consider:

$$\begin{gathered} d(f,h) = \inf \Big\{ \varepsilon > 0 : \text{there exists} \\ \lambda \in \Lambda_\varepsilon \text{ for which } \sup_{z \in [0,c)^2} |f(\lambda(z)) - h(z)| < \varepsilon \Big\}. \end{gathered} \tag{2.E.4}$$

It can be shown that $d(\cdot)$ is, in fact, a metric, and that equipping $D[0,c)^2$ with $d(\cdot)$ turns it into a Polish space (see Chapter 3 in Billingsley, 1999, for a related discussion of the construction of a metric for $D[0,1]$, and Straf, 1972, for its extension to more general spaces including the space of interest in this paper).

2.E.3 Probability Distributions

The metric defined above is used to construct the measurable space $(D[0,c)^2, \mathscr{D})$, in which $\mathscr{D}$ denotes the Borel sets associated with the Skorokhod Polish space $D[0,c)^2$. It is left to check whether it is possible to define a sufficiently large set of probability distributions on this space. Consider a field X defined on a discrete space, such as $\mathbb{N}^2$. It is well-known that, by Kolmogorov's Theorem, a probability distribution can be defined and characterized by the finite-dimensional distributions of $X(z_1), \ldots, X(z_n)$, for each choice of n and $z_1, \ldots, z_n \in \mathbb{N}^2$. This result does not hold for fields with continuous indices, such as those in $D[0,c)^2$. Some regularity conditions on the set of all finite-dimensional distributions, called *tightness conditions*, are required to define distributions on spaces of fields with continuous indices (see, for example, Section 1 titled "Measures on Metric Spaces" in Billingsley, 1999). These conditions are, in general, rather difficult to check, but it is known that we can define (i) Gaussian fields (with some conditions on the covariance operators), (ii) diffusion processes (with conditions on the instantaneous drift and volatility functions), and (iii) continuous (although possibly non-linear) transforms of these two objects and their distributions.

2.E.4 Convergence

Let us now fix a probability measure P_0 on $\mathscr{D}$. By equipping $(D[0,c)^2, \mathscr{D})$ with P_0, we turn it into a probability space. We can, therefore, define (i) almost sure convergence, (ii) convergence in probability, and (iii) weak convergence (i.e. convergence in distribution).[19] The following result is sometimes known

[19] The notions are respectively denoted (i) $X_n \xrightarrow{a.s.} X$, (ii) $X_n \xrightarrow{p} X$, and (iii) $X_n \xrightarrow{d} X$.

as the Skorokhod Theorem:

Theorem 2.2. Let (X_n) denote a sequence of random fields. If $X_n \overset{d}{\to} X$, then there exists another sequence of random fields (Y_n) such that (i) the distribution of Y_n is equal to the distribution of X_n, for each $n \in \mathbb{N}$, (iii) the distribution of Y is equal to the distribution of X, and (iv) $Y_n \overset{a.s.}{\to} Y$.

The Skorokhod Theorem relates the notions of almost sure convergence and weak convergence. It is a powerful tool for deriving some non-trivial results. Since every probability measure on the metric space $(D[0,c)^2, \mathscr{D})$ is the distribution of some random field, we know that, if a sequence of probability measures (P_n) converges to P, then there exists a sequence of random fields (X_n) such that (i) X_n has distribution P_n, for every $n \in \mathbb{N}$, (ii) X has distribution P, and (iii) $X_n \overset{d}{\to} X$. The Skorokhod Theorem says that we can construct this sequence (and its limit) on a common probability space and that we can construct it so that it converges *almost surely* to its limit.

2.E.5 The Continuous Mapping Theorem

The Continuous Mapping Theorem is a direct extension of the Skorokhod Theorem:[20]

Theorem 2.3. Consider a continuous (non-linear) transform $A(\cdot)$ from $(D[0,c)^2, \mathscr{D})$ to $(D[0,c)^2, \mathscr{D})$ such that $Z = A(X)$. If (X_n) satisfies $X_n \overset{d}{\to} X$, then $Z_n = A(X_n) \overset{d}{\to} Z$.

If we have a (non-linear) continuous transform $A(\cdot)$ that maps from $(D[0,c)^2, \mathscr{D})$ to $(D[0,c)^2, \mathscr{D})$, then every probability measure P_0 on $\mathscr{D}$ induces a probability measure Q_0 on $\mathscr{D}$, defined by $Q_0(S) = P_0(\text{inv}A(S))$, for every $S \in \mathscr{D}$, where $\text{inv}A(S)$ denotes the (measurable) set of fields $s \in D[0,c)^2$ for which $A(s) \in S$. The Continuous Mapping Theorem says that, if a sequence of probability measures (P_n) converges to P, then the sequence of probability measures (Q_n) induced by the elements of (P_n) converges to the probability measure Q induced by P. This result is also valid for bi-dimensional fields.

2.E.6 Dual Spaces, Differentiability, and the Delta Method

The aim of this section is to briefly explain how the delta method can be extended to be applied to random fields. I avoid certain topological details, for exposition. First, I discuss the Riesz-Markov-Kakutani theorems. Second, I relate measures with cadlag functions. Third, I use the theory of distributions (or generalized functions) to formalize the notion of a differential for our cadlag functions (see, for instance, Schwartz, 1966, and Gelfand and Vilenkin, 1964).

Riesz-Markov-Kakutani Theorems

The Riesz-Markov-Kakutani representation theorems consider spaces of continuous functions. Several spaces have been considered—for example, the space of continuous functions with compact support and

[20] See page 20 in Billingsley (1999).

the space of continuous functions that vanish at infinity. Let $\mathscr{C}$ denote one of these spaces. These theorems characterize the continuous linear functional on this space—that is, the elements of the dual space, say $\mathscr{C}^*$, of $\mathscr{C}$.

Theorem 2.4. Any continuous linear functional Ψ on $\mathscr{C}$ can be written as an integral:

$$\Psi(f) = \int_X f(x) d\mu(x) \equiv \langle f, \mu \rangle, \tag{2.E.5}$$

where μ is a measure. This measure is unique.

The set of measures associated with $\mathscr{C}^*$ depends on the definition of $\mathscr{C}$. For instance, when $\mathscr{C}$ is the space of continuous functions vanishing at infinity, $\mathscr{C}^*$ is the space of measures with bounded variation. This theorem implies that it is equivalent to know the unique measure μ or its impact $\langle f, \mu \rangle$ on every element f of the space $\mathscr{C}$.

Measures and Cumulative Functions

We can also analyze the impact of a measure on other functions. For instance, the cumulative function associated with a measure μ is defined by the following integral:

$$H(x) = \int_X \mathbb{1}_{u_1 < x_1, u_2 < x_2} d\mu(u) = \int_0^{x_1} \int_0^{x_2} d\mu(u), \tag{2.E.6}$$

for every $x \in X$. The cumulative function $H(\cdot)$ is cadlag with bounded variation, ensuring that it is equivalent to know the unique measure μ or the cumulative function $H(\cdot)$. This argument is the basis for the Riemann-Stieltjes integral, explaining why we often encounter the following notation in practice:

$$\langle f, \mu \rangle = \int_X f(x) dH(x) \equiv \langle f, H \rangle. \tag{2.E.7}$$

The trajectories of the random fields in Section 2.3 are cadlag with bounded variation. Therefore, each trajectory is associated with a unique measure μ. Since these trajectories depend on $\omega \in \Omega$, each random field is associated with a stochastic cumulative function, say $H(x; \omega)$. This setting induces the following stochastic integral:

$$\langle f, H(\cdot; \omega) \rangle = \int_X f(x) dH(x; \omega), \tag{2.E.8}$$

with ω varying.

Partial Derivatives

The notation $\int_x f(x) dH(x)$ gives the impression that $H(\cdot)$ is differentiable when it is not—of course, it may not even be continuous. Let us take a moment to understand this notation. First, notice that, if $\mathscr{C}_0 \subseteq \mathscr{C}_1$, then $\mathscr{C}_0^* \subseteq \mathscr{C}_1^*$. It is, therefore, natural to begin by considering a subset of $\mathscr{C}$. To this end, let $\mathscr{C}_{K,1}$ denote the subspace of continuously-differentiable functions with compact rectangular support $K = [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$. It can be shown that the measure μ—or equivalently, the cumulative function $H(\cdot)$—is characterized by its impact $\langle f, H\rangle$ on each function $f \in \mathscr{C}_{K,1}$. Now, let us consider the linear functional and integrate by parts with respect to x_2. This procedure yields:

$$\begin{aligned}\langle f, H\rangle &= \int_{\underline{x}_1}^{\bar{x}_1}\int_{\underline{x}_2}^{\bar{x}_2} f(x_1, x_2) H(dx_1, dx_2)\\ &= \int_{\underline{x}_1}^{\bar{x}_1}\left\{\Big[f(x_1, x_2)H(dx_1, x_2)\Big]_{\underline{x}_2}^{\bar{x}_2} - \int_{\underline{x}_2}^{\bar{x}_2}\frac{\partial f(x_1, x_2)}{\partial x_2} H(dx_1, x_2)dx_2\right\}\\ &= -\int_{\underline{x}_1}^{\bar{x}_1}\int_{\underline{x}_2}^{\bar{x}_2}\frac{\partial f(x_1, x_2)}{\partial x_2} H(dx_1, x_2)dx_2.\end{aligned} \tag{2.E.9}$$

The right-hand side of the final equality is a continuous functional of f and can be written as, say $\langle f, H_1\rangle$, by the representation theorem above, where H_1 is defined by:

$$\int_X f(x)H_2(dx) = -\int_X \frac{\partial f(x_1, x_2)}{\partial x_2} H(dx_1, x_2)dx_2. \tag{2.E.10}$$

If we compare this expression with the standard formula for integration by parts, we see that H_2 can be interpreted as the partial derivative of H with respect to x_2. We can, similarly, define its partial derivative with respect to x_1:

$$\int_X f(x)H_1(dx) = -\int_X \frac{\partial f(x_1, x_2)}{\partial x_1} H(x_1, dx_2)dx_1. \tag{2.E.11}$$

To summarize, the (partial) derivative of a cadlag function—that is, the derivative of a measure—is itself a measure. This derivative can be used in the usual way whenever we are concerned with its impact on continuously-differentiable functions. Moreover, the partial derivatives of H are stochastic whenever the measure of H is stochastic. We can, therefore, construct partial derivatives of our random fields whenever we need them to compute stochastic integrals.

2.F Prices and Quantities for Aggregate Goods

If there were only two goods (with precise characteristics and identical quantity units) and if their markets were competitive, then all consumers would face the same prices at each date t. In such an environment, we would only observe a small number of distinct prices, making it impossible to non-parametrically identify preferences. This feature of the identification problem motivates a literature on the set identification of preferences, starting with Varian (1982), who builds on work by Afriat (1967). This reasoning also underlies the construction of price indices with values independent of the individuals, but possibly depending on the province or state—for instance, only 5 distinct prices per year are available in the application in Dette et al. (2016). However, with scanner data, there often exists significant variation in prices and expenditures (see Chernozhukov et al., 2020):

(i) Prices almost always vary across retailers (even for homogeneous goods). This problem is inflated by sales, discounts, and wholesale pricing. For example, in the empirical application in Section 1.5, a twenty-four pack of Miller Lite ranges from \$6.99 to \$35.63. This wide range of prices exists, without even accounting for the fact that packs are, in general, although not always, cheaper per bottle.

(ii) To obtain a reasonable number of goods, we need to aggregate a large number of different homogeneous goods, without artificially forcing prices to be the same across consumers, as in Echenique et al. (2011), Dette et al. (2016), Kitamura and Stoye (2018), Deb et al. (2018), or Allen and Rehbeck (2019). This procedure yields even more price variation. For example, in our empirical application, the price of beer ranges from \$0.63 to \$21.98 per litre.

(iii) It is also common to observe lot of variation in expenditure. For example, in our empirical application, monthly expenditure ranges from \$6.23 to \$2767.76. It would be inappropriate to restrict all consumers to have the same expenditure, as in Kitamura and Stoye (2018).

2.F.1 Price and Quantity Indices

Assume that (aggregate) good j is composed of K_j homogeneous goods. In practice, K_j can be very large. For example, the Nielsen dataset contains information on "three million unique [universal product codes] for 1073 products in 106 product groups" (see Ng, 2017, Guha and Ng, 2019, and the detailed description in Section 2.6). In general, for each consumer i and date t, we observe a price p_{ijkt} and quantity x_{ijkt}, for every good $k = 1, \ldots, K_j$ and group $j = 1, 2$. We need to transform our observations into a common quantity unit, and define aggregate prices (p_{ijt}) and quantities (x_{ijt}). These amounts are aggregated over goods, but depend on the consumer, group, and month.

To aggregate goods, it is necessary to define a *benchmark*—typically, a representative consumer at a reference date. To this end, let us consider average expenditure in aggregate good j at reference date 0:

$$\bar{E}_{j0} = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K_j} p_{ijk0}x_{ijk0}. \tag{2.F.1}$$

Equivalently, we can write:

$$\bar{E}_{j0} = \sum_{k=1}^{K_j}\left\{\left(\frac{1}{n}\sum_{i=1}^{n} p_{ijk0}\right)\left(\sum_{i=1}^{n} p_{ijk0}x_{ijk0}\right)\left(\sum_{i=1}^{n} p_{ijk0}\right)^{-1}\right\} = \sum_{k=1}^{K_j} p_{jk0}x_{jk0}, \tag{2.F.2}$$

where p_{jk0} denotes the average price of the k^{th} homogeneous good in aggregate group j at reference date 0, and x_{jk0} denotes a weighted sum of quantities x_{ijk0}. Precisely:

$$p_{jk0} = \frac{1}{n}\sum_{i=1}^{n} p_{ijk0} \ \text{ and } \ x_{jk0} = \left(\sum_{i=1}^{n} p_{ijk0}x_{ijk0}\right)\left(\sum_{i=1}^{n} p_{ijk0}\right)^{-1}. \tag{2.F.3}$$

We can average prices because they concern homogeneous groups of goods.

Now, we can construct prices (p_{ijt}) and quantities (x_{ijt}) for each consumer i, group j, and date t with the Laspeyres and Paasche indices. To do so, we need to consider the ratio of expenditure $E_{ijt} = \sum_{k=1}^{K_j} p_{ijkt}x_{ijkt}$ and benchmark expenditure $\bar{E}_{j0}$. This ratio can be written as:

$$\frac{E_{ijt}}{\bar{E}_{j0}} = \left(\sum_{k=1}^{K_j} p_{ijkt}x_{ijkt}\right)\left(\sum_{k=1}^{K_j} p_{jk0}x_{jk0}\right)^{-1} = \mathcal{L}_{ijt}\mathcal{P}_{ijt}, \tag{2.F.4}$$

where $\mathcal{L}_{ijt}$ denotes the Laspeyres price index, defined by:

$$\mathcal{L}_{ijt} = \left(\sum_{k=1}^{K_j} p_{ijkt}x_{jk0}\right)\left(\sum_{k=1}^{K_j} p_{jk0}x_{jk0}\right)^{-1}, \tag{2.F.5}$$

and $\mathcal{P}_{ijt}$ denotes the Paasche quantity index, defined by:

$$\mathcal{P}_{ijt} = \left(\sum_{k=1}^{K_j} p_{ijkt}x_{ijkt}\right)\left(\sum_{k=1}^{K_j} p_{ijkt}x_{jk0}\right)^{-1}. \tag{2.F.6}$$

Intuitively, $\mathcal{L}_{ijt}$ is the (relative) evolution of the aggregate price between (p_{ijkt}, x_{ijkt}) and the benchmark. The interpretation is similar for $\mathcal{P}_{ijt}$. We can, as a result, write:

$$p_{ijt} = P_j\mathcal{L}_{ijt} \ \text{ and } \ x_{ijt} = X_j\mathcal{P}_{ijt}, \tag{2.F.7}$$

where P_j and X_j are values to be fixed.

By the coherency of the definition of expenditure in group j, we expect to obtain $\bar{E}_{j0} = P_jX_j$. There is, however, one remaining degree of freedom—the homogeneous goods in a group are usually measured in qualitatively dissimilar units (e.g. kilograms, litres, dozens), yielding a need for an Artificial Quantity Unit (AQU). Without loss of generality, I set an AQU by fixing $X_j = 1$. Under this restriction:

$$p_{ijt} = \sum_{k=1}^{K_j} p_{ijkt}x_{jk0} \ \text{ and } \ x_{ijt} = \left(\sum_{k=1}^{K_j} p_{ijkt}x_{ijkt}\right)\left(\sum_{k=1}^{K_j} p_{ijkt}x_{jk0}\right)^{-1}, \tag{2.F.8}$$

for every consumer i, group j, and date t. It is worth mentioning that *I have used the Laspeyres-Paasche decomposition in a non-standard way.* Indeed, it is usually applied to measure inflation and aggregated over both *goods and consumers*. I only aggregate over goods in order to keep the heterogeneity in prices at the consumer level.

2.F.2 Direct Choice of a Common AQU

The standard aggregation method above is simplified if we consider goods at a sufficiently disaggregate level. In order to illustrate, let us consider two aggregate groups:

Group 1: Beverages with an alcohol percentage between 3% and 8%;

Group 2: Beverages with an alcohol percentage between 9% and 15%.

The first group includes beers and ciders, and the second group includes wines. The separability assumption is almost satisfied between these groups and all other goods, although there can be some complementarity with food—for instance, the joint consumption of Sauternes and foie gras. In this example, it is easy to define an AQU—we can simply define the price p_{ijt} to be the ratio of expenditure E_{ijt} and the volume of all beverages, or volume of all alcohol, purchased by consumer i, in group j, at date t. It should be noted that preferences and aggregate data on groups of goods depend on the AQU.

If there is a lot of variability in the price of a unit within an aggregate group, then there will be a lot of variability in the price of that group. In general, we will not observe some very small or large prices, but, as we see in the application in Section 2.6, the observed prices can be well-distributed in a wide interval. Similar conclusions hold for expenditure. From now on, I assume that prices (p_{ijt}) and quantities (x_{ijt}) have been constructed using one of these approaches from the crude observable data.

Remark 2.8. Of course, in some datasets, we do not observe p_{ijkt} if the k^{th} good in group j is not purchased by consumer i at date t. This type of partial observability is problematic if we want to aggregate using the Laspeyres and Paasche indices, but not if we have a common AQU. See Crawford and Polisson (2016) and Section 5 in Chernozhukov et al. (2020).

2.G Regularizing the Nadaraya-Watson Estimator

The Nadaraya-Watson estimator is a local estimator, creating difficulty when deriving its asymptotic distributional properties when it is considered as a process indexed by z. It is possible to partially solve this difficulty by "projecting" this estimator onto a *partition of unity*. Projecting onto a partition of unity turns the local estimator into a global estimator. In this section, I discuss the approach in Zinde-Walsh (2018).

In Zinde-Walsh (2018), the estimator of interest does not necessarily have a closed-form solution, but it can be defined by means of integration. Let:

$$\hat{F}(x|z) = \frac{\sum_{i=1}^{n} \mathbb{1}\{x_i \leq x\} K\left(\frac{z_i - z}{h}\right)}{K\left(\frac{z_i - z}{h}\right)} = \frac{\frac{1}{nh}\sum_{i=1}^{n} \mathbb{1}\{x_i \leq x\} K\left(\frac{z_i - z}{h}\right)}{\hat{f}(x)}, \tag{2.G.1}$$

denote the usual kernel estimator for a conditional cumulative distribution function, obtained by integrating the kernel estimator for the conditional density.[21] If $\varphi(\cdot)$ is a continuously-differentiable function of x with bounded support, then, by integration by parts, we obtain:

$$\int \varphi(x)\hat{f}(x|z)dx = -\int \hat{F}(x|z)\frac{d\varphi(x)}{dx}dx. \tag{2.G.2}$$

Of course, this simple formula fails to hold when $\varphi(\cdot)$ has unbounded support. Therefore, if we are interested in the expectation of a function of x with unbounded support, as with the conditional mean for which $\varphi(x) = x$, then we cannot use this relationship.

Let us now show how we can use a *partition of unity*. A partition of unity on $\mathbb{R}_+$ is a set $\mathcal{R}$ of functions $\rho : \mathbb{R}_+ \to [0,1]$ that satisfy (i) $\sum_{\rho\in\mathcal{R}} \rho(x) = 1$, at every $x \geq 0$, and (ii) for every $x \geq 0$, there exists a neighbourhood of x on which a finite number of functions in $\mathcal{R}$ are non-zero. For example:

$$\rho_n(x) = \begin{cases} x-(n-1), & \text{if } n-1 < x < n, \\ 1-(x-n), & \text{if } n < x < n+1, \\ 0, & \text{otherwise,} \end{cases} \tag{2.G.3}$$

for every $x \geq 0$ and each $n \in \mathbb{N}$, is a partition of unity on $\mathbb{R}_+$. Indeed, exactly two functions in this set are positive at each point in $\mathbb{R}_+$ and these two functions sum up to one. I illustrate several elements of this partition of unity in Figure 2.17(a).

Here, we will require a partition of unity consisting of continuously-differentiable functions with bounded support. There are many ways to construct a partition with these properties (see Christensen and Goh, 2017). We can use the partition of unity:

$$\rho_n(x|\gamma) = \begin{cases} \frac{1}{2}\Big\{1-\cos\big[\pi(\gamma x - n)\big]\Big\}, & \text{if } \frac{n}{\gamma} < x < \frac{n+2}{\gamma}, \\ 0, & \text{otherwise,} \end{cases} \tag{2.G.4}$$

for every $x \geq 0$ and each $n \in \mathbb{N}$, in which $\gamma > 0$ denotes a bandwidth parameter and π denotes the mathematical constant, rather than the density $\pi(\cdot)$ in Table 2.1. Once again, exactly two functions in this set are positive at each point in $\mathbb{R}_+$ and these two functions sum up to one. I illustrate several elements of this partition of unity in Figure 2.17(b).

Under a regularity condition on $\varphi(\cdot)$, we can apply Fubini's Theorem to construct a "smoothed"

[21] An extra condition on the support of $K(\cdot)$ needs to be introduced for this estimator to be valid.

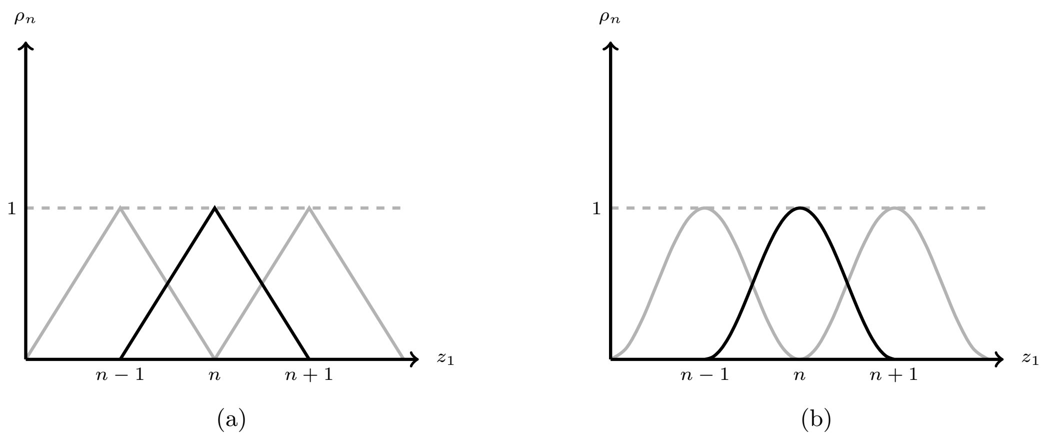

Figure 2.19. On the left, I illustrate ρ_{n-1}, ρ_n, and ρ_{n+1} in (2.G.3). On the right, I illustrate ρ_{n-1}, ρ_n, and ρ_{n+1} in (2.G.4). Dark lines highlight ρ_n.

kernel estimator $\tilde{f}(x|z)$ for the conditional density $f(x|z)$, defined by its impact on functions:

$$\begin{aligned} \int \varphi(x)\tilde{f}(x|z)dx &\equiv \sum_n \int \varphi(x)\rho_n(x)\tilde{f}(x|z)dx \\ &= -\sum_n \int \hat{F}(x|z)\frac{d}{dx}\left[\varphi(x)\rho_n(x)\right]dx, \end{aligned} \tag{2.G.5}$$

where the first relationship follows from Fubini's Theorem (see Assumption 4 in Zinde-Walsh, 2018). A smoothed Nadaraya-Watson estimator can be obtained by inputing $\varphi(x) = x$. The resulting estimator has the form:

$$\tilde{x}(z) = -\sum_n \int \hat{F}(x|z)\frac{d}{dx}\left[x\rho_n(x)\right]dx, \tag{2.G.6}$$

This slight change in the definition yields convergence (at a parametric rate) and lets us apply a functional version of the Central Limit Theorem.

A similar technique can be used to smooth the Nadaraya-Watson estimator with respect to the conditioning variable z. For this discussion, let us consider the Nadaraya-Watson estimator $\tilde{x}(z)$ defined in (2.G.6). This estimator is non-negative at each z. We can, thus, define the cumulative function:

$$\tilde{G}(z) = \int_0^z \tilde{x}(u)du, \tag{2.G.7}$$

and define a smoothed version $\tilde{\tilde{x}}(z)$ of the Nadaraya-Watson estimator by means of its impact on functions of z with bounded support. Again, the purpose of this procedure is to transform the local estimator into a global estimator. Without a global estimator, we cannot formally discuss notions like inversion. We could, instead, start with a global estimator, but the Nadaraya-Watson estimator is tractable, and this regularization will let us apply the theory of generalized functions to test the integrability of expected demand with a *parametric rate* (see Section 2.5.4).

2.H An Algorithm to Invert a Deterministic Field

The Nadaraya-Watson estimator can be used to invert a known deterministic function $A(\cdot)$. Indeed, the consistency of the Nadaraya-Watson estimator requires the sample distribution of $(z_j : j = 1, \ldots, J)$ to converge to a continuous distribution with support $\mathcal{Z}$, and this convergence can be reached with random z_j's, as assumed in Assumption 2.5, as well as with a "grid" of values for which the empirical distribution of the grid converges to a continuous limiting distribution π (see Gasser and Müller, 1984).

We can invert by applying the following steps:

Step 1: Construct a deterministic grid $(z_1, \ldots, z_J)$ for z.

Step 2: Compute $x_j = A(z_j)$, for each j.

Step 3: Use the joint observations to compute the Nadaraya-Watson estimator for $\mathbb{E}[Z|X = x]$. This estimator provides a consistent approximation of $\text{inv}A(x)$. This approach transforms the "irregular" grid $x_1, \ldots, x_J$ into a regular grid.

Chapter 3

Non-Parametric Taste Uncertainty

In Chapter 2, I use random fields to construct a model of consumption for scanner data with infinite-dimensional heterogeneity, and use this model to identify and estimate demand and preferences. However, there are many ways to construct an identified model with infinite-dimensional heterogeneity.

In this chapter, I introduce two models of non-parametric random utility for demand systems: the stochastic absolute risk aversion (SARA) model, and the stochastic safety-first (SSF) model. In each model, individual-level heterogeneity is characterized by a distribution $\pi \in \Pi$ of taste parameters, and heterogeneity across consumers is introduced using a distribution F over the distributions in Π. Demand is non-separable and heterogeneity is infinite-dimensional. Both models admit corner solutions. I consider two frameworks for estimation: a Bayesian framework in which F is known, and a hyperparametric (or empirical Bayesian) framework in which F is a member of a known parametric family. These methods are illustrated by an application to a large U.S. panel of scanner data on alcohol consumption.

3.1 Introduction

As described in the previous chapter, the recent availability of databases containing all dated purchases made by a large number of consumers (28,036 in the application) presents a modern challenge for the econometrics of demand systems, requiring new models and estimation approaches (see, for example, Burda et al., 2008, 2012, for discrete choice, and Guha and Ng, 2019, Chernozhukov et al., 2020, and Chapter 2 for the first analyses of such data in the demand literature). This type of data is commonly called *scanner data* because its collection involves retailers or households scanning each purchased good on the date of purchase. This chapter introduces two new models of random utility for scanner data: the stochastic absolute risk aversion (SARA) model, and the stochastic safety-first (SSF) model. These models have the following advantages in comparison with the existing literature:

(i) Both models are *consistent with consumer theory*: Every consumer maximizes a strictly increasing and strictly quasi-concave utility function. The latter property is not accommodated by existing approximations of the utility function like the quadratic approximation of the utility function (Theil and Neudecker, 1958; Barten, 1968), the translog utility model (Johansen, 1969; Christensen et al., 1975), or the Almost Ideal Demand System (Deaton and Muellbauer, 1980a) and its extensions (Banks et al., 1997; Moschini, 1998).

(ii) Both models are non-parametric. In each model, the utility function is indexed by a functional parameter characterizing the individual heterogeneity, allowing for *infinite-dimensional heterogeneity*. In this respect, this chapter differs from the existing literature when finite-dimensional heterogeneity is considered (see Beckert and Blundell, 2008, Blomquist et al., 2015, Blundell, Horowitz, and Parey, 2017, and Blundell, Kristensen, and Matzkin, 2017, for some examples of finite-dimensional restrictions). The non-parametric approach in this chapter is in line with Dette et al. (2016) who write, "in general the multivariate demand function is a non-monotonic function of an infinite-dimensional unobservable—the individual's preference ordering."

(iii) Both models yield demand functions with *non-separable heterogeneity* (see the discussions in Brown and Walker, 1989, Beckert and Blundell, 2008, and Dette et al., 2016). They are also endowed with precise *structural interpretations*, as heterogeneity is introduced by means of a distribution π of taste parameters, so that we can imagine consumers facing *taste uncertainty*, which they eliminate using expected utility.

(iv) Both models are *identified* under weak restrictions. Identification follows from the use of panel data. Without such data, we lose identification (Hausman and Newey, 2016b). Of course, the structure of scanner data is extremely important.

Each model is characterized by a basis of functions. This basis is used to generate a family of utility functions. A distribution is, then, placed over this family. To be precise, I start with a basis of increasing and concave functions. Let $U(x; a)$ denote an element of this basis, where x is a bundle and $a \in \mathscr{A}$ is a finite-dimensional vector of taste parameters. A family of utility functions is generated by taking the convex hull of the basis. Let $U(x; \pi) = \mathbb{E}_\pi\big[U(x; a)\big]$ denote an element of this family, where $\pi \in \Pi$ is a distribution on $\mathscr{A}$. This family is indexed by a functional parameter π, which can be structurally interpreted as taste uncertainty (resolved after the consumer makes her decisions). The heterogeneity across consumers is introduced using a distribution F on the set Π of probability distributions π on $\mathscr{A}$. Therefore, each model combines uncertainty and heterogeneity: the uncertainty in taste for a given consumer is represented by π, and the heterogeneity across consumers is captured by F.

The paper considers a two-good framework with continuous support for x. It is organized as follows: Section 3.2 introduces the stochastic absolute risk aversion (SARA) model and Section 3.3 introduces the stochastic safety-first (SSF) model. For each model, I derive conditions on Π under which there exists a unique demand system, for each $\pi \in \Pi$. In Section 3.4, the distribution of heterogeneity F

is introduced. When F is known, we obtain a Bayesian framework in which the functional parameter $\pi \in \Pi$ has to be estimated. When F is a member of a known parametric family, indexed by θ, we obtain an empirical Bayesian framework with a hyperparameter θ that has to be estimated, and a stochastic functional parameter π that has to be filtered. In Section 3.5, I consider the identification of the taste distribution π *within* each model. Next, I examine if it is possible to distinguish *between* stochastic risk aversion and stochastic safety-first. In Section 3.6, I use the Nielsen Homescan Consumer Panel to illustrate our methodology in an application to the consumption of alcohol. Section 3.7 concludes. The details of the Dirichlet process are in Appendix 3.A; integrability is discussed in Appendix 3.B; an optimization procedure for filtering the taste distributions π after estimating F is in Appendix 3.C; details of the data are placed in Appendix 3.D.

3.2 A Model with Stochastic Risk Aversion

This section introduces the first utility specification that is considered. It first describes the set of utility functions, then derives conditions under which there exists a unique demand system. The taste uncertainty is introduced using risk aversion parameters.

3.2.1 The Set of Utility Functions

There are two goods, denoted 1 and 2. Let $\bar{R} = \mathbb{R}^2_+$ denote the non-negative orthant with interior R. A consumer has preferences over the bundles in $\bar{R}$. Her preferences are summarized by a utility function of the form:

$$U(x;\pi) = -\mathbb{E}_\pi\big[\exp(-A'x)\big], \tag{3.2.1}$$

for every x such that $x_1, x_2 \geq 0$, where $A = (A_1, A_2)$ is a positive stochastic parameter characterizing the consumer's degrees of absolute risk aversion with respect to goods 1 and 2, and π is a joint distribution for this pair of stochastic taste parameters. Her preferences are, as a result, contained in a broad family of utility functions, indexed by a functional parameter π. There are two interpretations of specification (3.2.1): (i) the preferences are summarized by a deterministic utility function in the convex hull generated by a parametric family, or (ii) the consumer faces "taste uncertainty" and she resolves this uncertainty by using expected utility. These preferences will be referred to as *stochastic absolute risk aversion* (SARA) preferences.[1]

If π is a point mass at $a = (a_1, a_2)$ such that $a_1, a_2 > 0$, the stochastic parameters are constant, and

[1]These preferences differ from those used to describe consumer behaviour when facing ambiguity or uncertainty, as in, say, Halevy and Feltkamp (2005).

$U(x;\pi)$ reduces to $U(x;a) = -\exp(-a'x)$. This function is strictly increasing because we have:[2]

$$\frac{\partial U(x;a)}{\partial x} = \begin{bmatrix} a_1 \exp(-a'x) \\ a_2 \exp(-a'x) \end{bmatrix} > 0, \tag{3.2.2}$$

at each x such that $x_1, x_2 > 0$, and concave (although not necessarily strictly concave) because the Hessian associated with the utility function:

$$\frac{\partial^2 U(x;a)}{\partial x \partial x'} = -\exp(-a'x) \begin{pmatrix} a_1^2 & a_1 a_2 \\ a_1 a_2 & a_2^2 \end{pmatrix}, \tag{3.2.3}$$

is negative semi-definite, at each x such that $x_1, x_2 > 0$. This matrix is related to a bivariate measure of absolute risk aversion[3] (Richard, 1975; Karni, 1979, 1983; Grant, 1995). These properties translate into properties of the more general utility function: $U(x;\pi)$.

Proposition 3.1. If preferences are SARA and the consumer's taste distribution π is not the mixture of point masses at $a, a' \in R$ where a is proportional to a', then the utility function $U(x;\pi)$ is strictly increasing with a negative definite Hessian everywhere on R.

Proof. The utility function $U(x;\pi)$ is strictly increasing on R because:

$$\frac{\partial}{\partial x}\mathbb{E}_\pi \left[U(x;A)\right] = \mathbb{E}_\pi \left[\frac{\partial U(x;A)}{\partial x}\right] > 0, \tag{3.2.4}$$

at every x such that $x_1, x_2 > 0$. Its Hessian is negative definite on R because the sum of two 2-by-2 matrices of rank 1, whose columns are not proportional, has full rank. □

Proposition 3.1 implies that we have effectively constructed a family of well-behaved utility functions $\{U(x;\pi) : \pi \in \Pi\}$ indexed by a functional parameter π, describing the taste uncertainty, instead of the standard finite-dimensional parameter that is usually considered in the literature.

Let $g_\pi(\cdot)$ denote the function defined by the implicit equation:

$$U(x_1, g_\pi(x_1, u); \pi) = u, \tag{3.2.5}$$

for every $x_1 \geq 0$, and each (attainable) level of utility $u < 0$. This implicit equation has a unique solution because $U(x;\pi)$ is strictly increasing on $\bar{R}$. The function $g_\pi(\cdot, u)$ is the *indifference curve* associated with the functional parameter π and a utility level of u—$g_\pi(\cdot, u)$ maps every value of x_1 to a value of x_2

[2]Here, >0 means each component is strictly larger than 0.

[3]Such a measure can be defined as:

$$-\left(\operatorname{diag}\frac{\partial U(x;\pi)}{\partial x}\right)^{-1/2} \frac{\partial^2 U(x;a)}{\partial x \partial x'} \left(\operatorname{diag}\frac{\partial U(x;\pi)}{\partial x}\right)^{-1/2},$$

where $\operatorname{diag}\frac{\partial U(x;\pi)}{\partial x}$ is the diagonal matrix whose diagonal elements are the first derivatives of $U(x;\pi)$.

for which (x_1, x_2) attains a utility level of u given π. The implicit function theorem implies that $g_\pi(\cdot)$ is twice-continuously-differentiable with respect to x_1 and:

$$\frac{\partial g_\pi(x_1, v)}{\partial x_1} = -\text{MRS}(x_1, g_\pi(x_1, v); \pi), \tag{3.2.6}$$

on R where $\text{MRS}(x;\pi) \equiv \frac{\partial U(x;\pi)/\partial x_1}{\partial U(x;\pi)/\partial x_2}$ denotes the *marginal rate of substitution* at x—the rate at which the consumer is willing to exchange good 1 for good 2 given x and π. The indifference curve $g_\pi(\cdot, v)$ is strictly convex such that:

$$\frac{\partial^2 g_\pi(x_1, v)}{\partial x_1^2} > 0, \tag{3.2.7}$$

at every $x_1 > 0$, since the Hessian of $U(x;\pi)$ is negative definite everywhere on R (see Lemma 1 in Chapter 2). This property is stronger than the standard assumption of strict quasi-concavity, which allows this derivative to be zero on a nowhere dense set (Katzner, 1968). This distinction is important for what follows.

Note that, after integrating out the taste uncertainty, the absolute risk aversions will depend on the consumption level. For instance, when A_1 and A_2 are independent with distributions π_1 and π_2, the risk aversion for good 1 becomes:

$$A_1(x_1) = -\frac{d^2 U_1(x_1;\pi_1)/dx_1^2}{dU_1(x_1;\pi_1)/dx_1} = \frac{\mathbb{E}_{\pi_1}[A_1^2 \exp(-A_1 x_1)]}{\mathbb{E}_{\pi_1}[A_1 \exp(-A_1 x_1)]}, \tag{3.2.8}$$

where $U_1(x_1;\pi_1)$ denotes $\mathbb{E}_{\pi_1}[\exp(-A_1 x_1)]$, the portion of the utility function $U(x;\pi)$ corresponding to good 1. Clearly, $A_1(x_1)$ depends on x_1, as it is the average of A_1 given the following modified density:

$$\frac{A_1 \exp(-A_1 x_1)}{\mathbb{E}_{\pi_1}[A_1 \exp(-A_1 x_1)]}, \tag{3.2.9}$$

with respect to π_1.

3.2.2 The Demand Function

Let $z \in R$ denote a pair $z = (y, p)$ in which y denotes expenditure and p denotes the price of good 1, both normalized by the price of good 2. The consumer can purchase a bundle $x \in \bar{R}$ if, and only if, $px_1 + x_2 \leq y$. She chooses a bundle $x \in \bar{R}$ that solves:

$$\max_{x \in \bar{R}} U(x;\pi) \text{ subject to } px_1 + x_2 \leq y. \tag{3.2.10}$$

Let $X^*(z;\pi)$ denote the solution to:

$$\max_{x \in \mathbb{R}^2} -\mathbb{E}_\pi\big[\exp(-A'x)\big] \text{ subject to } px_1 + x_2 \leq y. \tag{3.2.11}$$

While (3.2.10) is restricted to bundles in the non-negative orthant, (3.2.11) allows for negative values. The solution to (3.2.11) is characterized by the following system of first-order conditions:

$$\mathrm{MRS}(x;\pi) \equiv \frac{\mathbb{E}_\pi\big[A_1 \exp(-A'x)\big]}{\mathbb{E}_\pi\big[A_2 \exp(-A'x)\big]} = p \;\text{ and }\; px_1 + x_2 - y = 0. \tag{3.2.12}$$

The first equality says that the marginal rate of substitution equals the relative price p. The second equality implies that the budget constraint holds with equality. Equivalently, we can solve the equality:

$$\mathbb{E}_\pi\big[(A_1 - pA_2)\exp(-(A_1 - pA_2)x_1)\exp(-A_2 y)\big] = 0, \tag{3.2.13}$$

for the first component $X_1^*(z;\pi)$, and then use the budget constraint in (3.2.12) to solve for $X_2^*(z;\pi)$. As long as $A_1 - pA_2$ is not almost surely equal to zero, the first-order partial derivative of the left side of this equality with respect to x_1 is strictly negative:

$$-\,\mathbb{E}_\pi\big[(A_1 - pA_2)^2\exp(-(A_1 - pA_2)x_1)\exp(-A_2 y)\big] < 0. \tag{3.2.14}$$

The function on the left side of (3.2.13) is, therefore, strictly decreasing in x_1, implying that there exists a unique solution $X_1^*(z;\pi)$ to (3.2.13), and a unique solution $X^*(z;\pi)$ to (3.2.11). If $X^*(z;\pi)$ is in $\bar{R}$, then $X^*(z;\pi)$ coincides with the solution to (3.2.10). Else, the solution to (3.2.10) is on the boundary of $\bar{R}$. Let $X(z;\pi)$ denote the solution to (3.2.10) given both z and π. There are three *regimes* of demand in the design space:

$$X(z;\pi) = \begin{cases} (0, y)', & \text{if } X_1^*(z;\pi) \le 0, \\ X^*(z;\pi), & \text{if } 0 \le X_1^*(z;\pi) \le y/p, \\ (y/p, 0)', & \text{if } y/p \le X_1^*(z;\pi). \end{cases} \tag{3.2.15}$$

Because the utility function $U(x;\pi)$ has strictly convex indifference curves everywhere on R, the demand function $X(z;\pi)$ is invertible in the second regime (see Proposition 2 in Chapter 2).

Proposition 3.2. If preferences are SARA and the consumer's taste distribution π is not the mixture of point masses at $a, a' \in R$ where a is proportional to a', then there exists a unique solution $X(z;\pi)$ to the maximization problem in (3.2.10) given z and π, for every $z \in R$, almost surely, for every π. There are three regimes of demand defined by (3.2.15). The resulting demand function $X(z;\pi)$ is invertible in the second regime.

To make a final remark, let us consider a *risk-neutral* consumer. In particular, let us assume that A_1 and A_2 tend stochastically to zero, with means that they tend to zero so that $\mathbb{E}_\pi[A_1]/\mathbb{E}_\pi[A_2]$ converges to a non-degenerate a_0. By considering the Taylor expansion of utility, it can be shown that these preferences are represented by:

$$x_1 + \frac{x_2}{a_0}. \tag{3.2.16}$$

This representation is unique up to an increasing transformation. For this risk-neutral consumer, goods are considered to be *perfect substitutes*. It is known that such a consumer will consume only good 1 whenever $p < a_0$, and only good 2 whenever $p > a_0$.

3.2.3 Gamma Taste Uncertainty

As an illustration, let us assume that A_1 and A_2 are independent and that A_j has a Gamma distribution $\gamma(\nu_j, \alpha_j)$ with *degree of freedom* $\nu_j > 0$ and *scale factor* $\alpha_j > 0$, for $j = 1, 2$. Under this specification, $\pi = \gamma(\nu_1, \alpha_1) \otimes \gamma(\nu_2, \alpha_2)$, where $\otimes$ denotes the tensor product of distributions. By the Laplace transform of the Gamma distribution:

$$U(x;\pi) = -\left(\frac{\alpha_1}{\alpha_1 + x_1}\right)^{\nu_1} \left(\frac{\alpha_2}{\alpha_2 + x_2}\right)^{\nu_2}. \tag{3.2.17}$$

Under this specification, the absolute risk aversion for good 1 in (3.2.8) becomes:

$$A_1(x_1) = \frac{1 + \nu_1}{\alpha_1 + x_1}, \tag{3.2.18}$$

which is hyperbolic in x_1. The indifference curve $g_\pi(\cdot)$ associated with utility level v is:

$$x_2 = g_\pi(x_1, v) \equiv \alpha_2 \left\{ \left[-\frac{1}{v} \left(\frac{\alpha_1}{\alpha_1 + x_1} \right)^{\nu_1} \right]^{\frac{1}{\nu_2}} - 1 \right\}, \tag{3.2.19}$$

for every $x_1 \geq 0$ and $v \in (-1, 0)$ such that:

$$x_1 < \alpha_1 \left[\left(-\frac{1}{v} \right)^{\frac{1}{\nu_1}} - 1 \right]. \tag{3.2.20}$$

It is easily shown that the second derivative of the indifference curve $g_\pi(\cdot, v)$ equals:

$$\frac{d^2 g_\pi(x_1, v)}{dx_1^2} = c \left(\frac{\alpha_1}{\alpha_1 + x_1} \right)^{\frac{\nu_1}{\nu_2} + 2} > 0, \tag{3.2.21}$$

for some $c > 0$. This inequality confirms that the indifference curve $g_\pi(\cdot, v)$ is strictly convex. Furthermore, the MRS is equal to:

$$\mathrm{MRS}(x;\pi) = \frac{\nu_1}{\nu_2} \frac{\alpha_2 + x_2}{\alpha_1 + x_1}. \tag{3.2.22}$$

The unconstrained solution $X_1^*(z;\pi)$ to the first-order condition in (3.2.12) is equal to:

$$X_1^*(z;\pi) = \frac{\nu_1}{\nu_1 + \nu_2} \cdot \frac{y}{p} + \frac{\nu_1 \alpha_2}{\nu_1 + \nu_2} \cdot \frac{1}{p} - \frac{\nu_2 \alpha_1}{\nu_1 + \nu_2}. \tag{3.2.23}$$

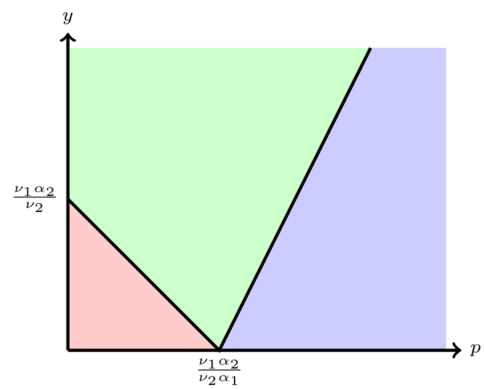

Figure 3.1. **Regimes for Gamma Taste Uncertainty.** The red region contains all designs z for which $X_1(z;\pi) > 0$ and $X_2(z;\pi) = 0$; the blue region contains all designs z for which $X_1(z;\pi) = 0$ and $X_2(z;\pi) > 0$; the green region contains all designs z for which $X_1(z;\pi) > 0$ and $X_2(z;\pi) > 0$.

The second component $X_2^*(z;\pi)$ is deduced from the budget constraint in (3.2.12). By equation (3.2.15), the demand function $X(z;\pi)$ coincides with $X^*(z;\pi)$ over the set $\mathcal{Z}$ of pairs z such that:

$$\min\left\{\nu_1 y - p\nu_2\alpha_1 + \nu_1\alpha_2,\ \nu_2 y + p\nu_2\alpha_1 - \nu_1\alpha_2\right\} > 0. \tag{3.2.24}$$

The three regimes of demand are illustrated in Figure 3.1 in the design space. The strict convexity of the indifference curve $g_\pi(\cdot, v)$ on $\mathcal{Z}$ implies that the demand function $X(\cdot;\pi)$ associated with this utility function is invertible on $\mathcal{Z}$.

3.3 A Model with Stochastic Safety-First

We now consider a model with taste parameters that have a *safety-first* interpretation.

3.3.1 The Set of Utility Functions

In Section 3.2, I constructed a family of well-behaved utility functions by taking the convex hull generated by a particular basis. In this section, I consider a second basis, consisting of functions with the form:

$$U(x;a) = (x_1 + a_1x_2) - (x_1 + a_1x_2 - a_2)^+ = \min\left\{x_1 + a_1x_2,\ a_2\right\}, \tag{3.3.1}$$

for every $x_1, x_2 \geq 0$, where $x^+ = \max\{0, x\}$ and $a_1, a_2 > 0$. This function corresponds to the "safety-first" criterion, introduced into the literature on portfolio management by Roy (1952). In order to illustrate, let us consider the consumption of alcohol, as in Chapter 2. Suppose that there are two groups of goods: group 1 consisting of drinks with low alcohol by volume such as beers and ciders, and group 2 consisting of drinks with high alcohol by volume such as wines and liquors. Assume that the quantities are measured in identical units such as volume of alcohol—that is, the total volume of the drink in litres multiplied by the alcohol by volume of the drink.[4] We can, then, add these volumes to aggregate two drinks with different sizes and/or percentages of alcohol. Here, a_1 is the consumer's relative preference between the two groups of drinks, and a_2 is a "control" parameter, specifying her attempt to limit her intake of alcohol.

Now, let us introduce a distribution π such that $\mathbb{E}_\pi[A_j] < \infty$, $j = 1, 2$, and define:

$$U(x;\pi) = \mathbb{E}_\pi\big[(x_1 + A_1 x_2) - (x_1 + A_1 x_2 - A_2)^+\big]. \tag{3.3.2}$$

By the law of iterated expectations, we obtain:

$$U(x;\pi) = x_1 + \mathbb{E}_\pi\big[A_1\big]x_2 - \mathbb{E}_\pi\mathbb{E}_\pi\big[(x_1 + A_1 x_2 - A_2)^+\big|A_1\big]. \tag{3.3.3}$$

These preferences are referred to as *stochastic safety-first* (SSF) preferences.

Under mild regularity conditions:

$$\begin{aligned}
\frac{\partial U(x;\pi)}{\partial x_1} &= 1 - \mathbb{E}_\pi\mathbb{E}_\pi\big[\mathbb{1}\{x_1 + A_1 x_2 - A_2 > 0\}\big|A_1\big], && (3.3.4)\\
&= \mathbb{E}_\pi\big[\mathbb{1}\{x_1 + A_1 x_2 - A_2 < 0\}\big], && (3.3.5)\\
\frac{\partial U(x;\pi)}{\partial x_2} &= \mathbb{E}_\pi\big[A_1\big] - \mathbb{E}_\pi\big[A_1\mathbb{E}_\pi\big[\mathbb{1}\{x_1 + A_1 x_2 - A_2 > 0\}\big|A_1\big]\big], && (3.3.6)\\
&= \mathbb{E}_\pi\big[A_1\mathbb{1}\{x_1 + A_1 x_2 - A_2 < 0\}\big], && (3.3.7)
\end{aligned}$$

for every x such that $x_1, x_2 > 0$. These partial derivatives are strictly positive when π has full support: $\pi(a_1, a_2) > 0$, for $a_1, a_2 > 0$. By taking the second-order derivatives:

$$\frac{\partial^2 U(x;\pi)}{\partial x \partial x'} = -\begin{pmatrix} \mathbb{E}_\pi[\pi_0] & \mathbb{E}_\pi[A_1\pi_0] \\ \mathbb{E}_\pi[A_1\pi_0] & \mathbb{E}_\pi[A_1^2\pi_0] \end{pmatrix}, \tag{3.3.8}$$

for every x such that $x_1, x_2 > 0$, where $\pi_0 \equiv \pi(x_1 + A_1 x_2|A_1)$ in which $\pi(\cdot|A_1)$ denotes the conditional density of A_2 given A_1, assuming that such a density exists. This matrix is both symmetric and negative definite when $\pi(\cdot|A_1)$ is continuous and A_1 is not constant. This result follows from the positivity of $\mathbb{E}_\pi[\pi_0]$ and the following equality:

$$\det\frac{\partial^2 U(x;\pi)}{\partial x \partial x'} = \mathbb{E}_\pi[\pi_0]V_{\tilde{\pi}}(A_1) > 0, \tag{3.3.9}$$

which holds for every x such that $x_1, x_2 > 0$, in which $\tilde{\pi}$ denotes the modified density:

$$\tilde{\pi}(a) = \frac{\pi(x_1 + a_2 x_2|a_1)\pi(a)}{\mathbb{E}_\pi[\pi_0]}. \tag{3.3.10}$$

[4] Quantities could be, alternatively, measured in calories.

Proposition 3.3. If preferences are SSF and the consumer's taste distribution π is continuous with full support given A_1, then the utility function $U(x;\pi)$ is strictly increasing with a negative definite Hessian everywhere on R.

Consequently, we have constructed another family of well-behaved utility functions $\{U(x;\pi) : \pi \in \Pi\}$ indexed by a functional parameter π, describing taste uncertainty.

3.3.2 The Demand Function

Let us revisit the utility maximization problem in (3.2.10). Under the safety-first specification, the analogue of the unconstrained first-order condition in (3.2.13) is given by:

$$\mathbb{E}_\pi[(1 - pA_1)\mathbb{1}\{x_1(1 - pA_1) + A_1 y - A_2 < 0\}] = 0. \tag{3.3.11}$$

I obtain this equality by equating the marginal rate of substitution with the relative price p, and then using the budget constraint to replace x_2 with $y - px_1$. Under the regularity conditions from above, the left-hand side is strictly monotone in x_1 given π, so that there exists a unique solution to the first-order condition. As in Section 3.2, I let $X_1^*(z;\pi)$ denote this solution, and let $X_2^*(z;\pi)$ denote the quantity $y - pX_1^*(z;\pi)$.

Proposition 3.4. If preferences are SSF and the consumer's taste distribution π is continuous with full support given A_1, then there exists a unique solution $X(z;\pi)$ to the maximization problem in (3.2.10) given z and π, for every $z \in R$, almost surely, for every π. There are three regimes of demand defined by (3.2.15). The resulting demand function $X(z;\pi)$ is invertible in the second regime.

When the consumer's preferences are SSF, the MRS has the form:

$$\mathrm{MRS}(x;\pi) \equiv \frac{\mathbb{E}_\pi[\mathbb{1}\{x_1 + A_1 x_2 - A_2 < 0\}]}{\mathbb{E}_\pi[A_1 \mathbb{1}\{x_1 + A_1 x_2 - A_2 < 0\}]} = \frac{1}{\mathbb{E}_\pi[A_1 | x_1 + A_1 x_2 - A_2 < 0]}. \tag{3.3.12}$$

Thus, the rate at which the consumer is willing to exchange good 1 for good 2 given x and π is equal to the inverse of the expectation of her relative preference between goods A_1, conditional on not surpassing her control parameter A_2.

Some functionals of distribution π can be especially interesting. For instance, in an application to the consumption of alcohol, we might expect the conditional distribution of A_2 given $A_1 = a_1$ to be concentrated around a single mode, characterizing an implicit alcohol limit for this consumer. Then, we can ask the following questions:

(i) Is this limit positively correlated with A_1? In other words, is there a positive relationship between this limit and a preference for *strong* alcoholic beverages?

(ii) Does a change in the maximum blood alcohol level for driving affect this limit?

These are questions that cannot be answered using classical demand systems like the Almost Ideal Demand System (Deaton and Muellbauer, 1980a). In fact, tests based on the Almost Ideal Demand System have rejected rationality in applications to alcohol consumption (Alley et al., 1992). Clearly, it is possible that the Almost Ideal Demand System is misspecified.

3.3.3 Exponential Threshold Taste Uncertainty

In general, the first-order condition in (3.3.11) has no closed-form solution. However, its expression can be simplified for some taste distributions π. As an illustration, let us assume that:

(i) A_1 and A_2 are independent.

(ii) A_2 has an exponential distribution $\gamma(1, \lambda)$ with survival function:

$$P(A_2 > a_2) = \exp(-\lambda a_2). \tag{3.3.13}$$

(iii) A_1 has a distribution with all of its moments and Laplace transform: $\Psi(v) = \mathbb{E}[\exp(-vA_1)]$, $v \geq 0$.

Under this specification, we can first integrate with respect to A_2 within the expectation in (3.3.11) in order to obtain the following condition:

$$\mathbb{E}_\pi\big[(1 - pA_1)\exp\{-\lambda(x_1 + (y - x_1 p)A_1\}\big] = 0. \tag{3.3.14}$$

Equivalently, we obtain:

$$\mathbb{E}_\pi\big[\exp\{-\lambda(y - x_1 p)A_1\}\big] - p\mathbb{E}_\pi\big[A_1 \exp\{-\lambda(y - x_1 p)A_1\}\big] = 0. \tag{3.3.15}$$

This equation can be written in terms of the Laplace transform Ψ for A_1. This yields:

$$\Psi\big[\lambda(y - x_1 p)\big] + p\frac{d\Psi}{dv}\big[\lambda(y - x_1 p)\big] = 0, \tag{3.3.16}$$

which can also be written as:

$$\frac{d\log\Psi}{dv}\big[\lambda(y - x_1 p)\big] = -\frac{1}{p}. \tag{3.3.17}$$

Finally, by inverting this expression and rearranging the terms, we get:

$$X_1^*(z;\pi) = \frac{1}{p}\left[y - \frac{1}{\lambda}\left(\frac{d\log\Psi}{dv}\right)^{-1}\left(-\frac{1}{p}\right)\right], \tag{3.3.18}$$

The second component $X_2^*(z;\pi)$ of the unconstrained solution in (3.3.11) is deduced from the budget constraint. It follows from equation (3.2.15) that the demand function $X(z;\pi)$ coincides with $X^*(z;\pi)$

if, and only if:

$$0 \leq \frac{1}{\lambda}\left(\frac{d\log\Psi}{dv}\right)^{-1}\left(-\frac{1}{p}\right) \leq y. \tag{3.3.19}$$

For instance, if A_1 follows a gamma distribution $\gamma(\nu,\alpha)$, then $\log\Psi(v) = -\nu\log(1+v/\alpha)$, and we obtain:

$$\frac{d\log\Psi(v)}{dv} = -\frac{\nu}{\alpha+v}, \tag{3.3.20}$$

for $v \geq 0$. Moreover, by inverting this function, we get:

$$\left(\frac{d\log\Psi}{dv}\right)^{-1}(\xi) = -\left(\frac{\nu}{\xi}+\alpha\right). \tag{3.3.21}$$

Therefore, the solution $X_1^*(z;\pi)$ has the form:

$$X_1^*(z;\pi) = \frac{1}{p}\left[y + \frac{1}{\lambda}(\alpha - \nu p)\right], \tag{3.3.22}$$

and demand $X(z;\pi)$ coincides with $X^*(z;\pi)$ if, and only if:

$$0 \leq \frac{\nu p - \alpha}{\lambda} \leq y. \tag{3.3.23}$$

The regimes of demand are illustrated in Figure 3.2 in the design space. We can also verify that the *Slutsky coefficient* is strictly negative[5] such that:

$$\Delta_x(z) \equiv \frac{\partial X_1(z;\pi)}{\partial p} + X_1(z;\pi)\frac{\partial X_1(z;\pi)}{\partial y} = -\frac{\nu}{\lambda p} < 0, \tag{3.3.24}$$

ensuring that the demand function $X(\cdot;\pi)$ is invertible over the set $\mathcal{Z}$ of pairs z on which demand is strictly positive (see Section 2.3 in Chapter 2).

3.4 Individual Heterogeneity

Sections 3.2 and 3.3 introduced two utility specifications, both indexed by the functional parameter π. Of course, different consumers can have different functional parameters. This individual heterogeneity is introduced in a second layer, by specifying a distribution F over the set Π of distributions on R, such as the Dirichlet process (see, for example, Navarro et al., 2006, for an application of the Dirichlet process in modelling individual differences). More precisely, I make the following assumption:

Assumption 3.1 (Latent Stochastic Model)**.**

(i) There are $n \geq 1$ consumers.

[5]This property holds for any Laplace transform Ψ of A_1 (see Appendix 3.B).

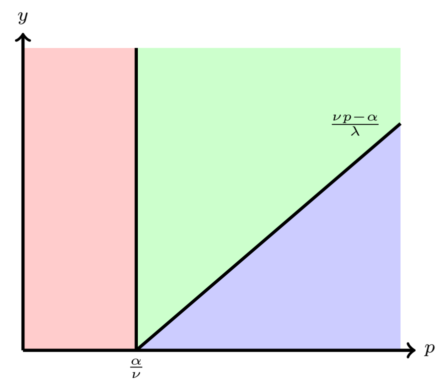

Figure 3.2. **Regimes for Exponential Threshold Taste Uncertainty.** The red region contains all designs z for which $X_1(z;\pi) > 0$ and $X_2(z;\pi) = 0$; the blue region contains all designs z for which $X_1(z;\pi) = 0$ and $X_2(z;\pi) > 0$; the green region contains all designs z for which $X_1(z;\pi) > 0$ and $X_2(z;\pi) > 0$.

(ii) Consumers are segmented into M homogeneous groups.

(iii) Consumers in group m have the utility function $U(x;\pi_m)$, for all $m = 1,\ldots,M$.

(iv) The taste parameters (π_m) are independently drawn from a Dirichlet process F.

Assumption 3.1 introduces a distribution F over the functional taste parameter π. This distribution F characterizes the heterogeneity across homogeneous groups. It can encompass, for example, regional or demographic differences in preferences. This infinite-dimensional heterogeneity is non-separable in the stochastic demand equation.

The Dirichlet process can be constructed in three steps:

Step 1: Consider the set of (Bernoulli) distributions on $\{0,1\}$. This set is characterized by $q \in \bar{R}$ such that $q_1 + q_2 = 1$. A distribution defined on this set of distributions is a distribution defined on this parameter set. We can, for instance, introduce a *beta distribution*, denoted $B(\alpha_1,\alpha_2)$. The distribution $B(\alpha_1,\alpha_2)$ has a continuous density:

$$f(q) = \frac{\Gamma(\alpha_1+\alpha_2)q_1^{\alpha_1}q_2^{\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}, \tag{3.4.1}$$

with respect to the Lebesgue measure over the simplex $\{(q_1,q_2) \geq 0 : q_1 + q_2 = 1\}$, where Γ denotes the gamma function,[6] and $\alpha_1, \alpha_2 > 0$ are positive scalar parameters.

Step 2: The beta distribution can be extended to define a distribution on the set of discrete distributions with weights $q_j \geq 0$, $j = 1,\ldots,J$, such that $\sum_{j=1}^{J} q_j = 1$. This procedure leads to the *Dirichlet*

[6]The gamma function Γ is defined by $\Gamma(\alpha) = \int_0^\infty \exp(-x)x^{\alpha-1}dx$, for each $\alpha > 0$.

distribution, denoted $D(\alpha)$. The resulting distribution $D(\alpha)$ has continuous density:

$$f(q) = \frac{\Gamma\left(\sum_{j=1}^{J} \alpha_j\right) \prod_{j=1}^{J} q_j^{\alpha_j}}{\prod_{j=1}^{J} \Gamma(\alpha_j)}, \tag{3.4.2}$$

with respect to the Lebesgue measure over the simplex:

$$\left\{ q \in \mathbb{R}_+^J : \sum_{j=1}^{J} q_j = 1 \text{ and } q_j \geq 0, \forall j \right\}, \tag{3.4.3}$$

(see, for example, Kotz et al., 2000, page 485, and Lin, 2016, for details).

Step 3: Then, the Dirichlet distribution can be extended to define a distribution on a large set of distributions[7] defined on $\bar{R}$ (see Appendix 3.A). This procedure leads to the *Dirichlet process*. The Dirichlet process is characterized by a distribution μ on $\bar{R}$ and a scaling parameter $c > 0$. The distribution μ can be thought of as the mean of the Dirichlet process, while the parameter c manages its degree of discretization (see Appendix 3.A). This extension of the Dirichlet distribution is much more complicated than the Dirichlet distribution, especially because the notion of the Lebesgue measure on the set of distributions, and the notion of a density, no longer exist (see Ferguson, 1974, Rolin, 1992, and Sethuraman, 1994).

Now, consider the implications of Assumption 3.1: If the functional and scaling parameters of the Dirichlet process are known, then we are in a Bayesian framework (see, for example, Geweke, 2012, for a Bayesian analysis of revealed preference) in which the mean of the posterior distribution of $\pi \in \Pi$ has to be estimated. Otherwise, we can assume that the mean μ of our process F is characterized by a finite-dimensional hyperparameter θ. Naturally, the hyperparametric model has two types of parameters: the hyperparameter θ to be estimated, and the functional parameters (π_m) to be filtered.

3.5 Non-Parametric Identification

In this section, I consider the identification of the functional parameter π *within* each model from the observation of a demand function, then whether we can distinguish *between* the SARA and SSF models.

Intuitively, a consumer's demand function is identified if we observe her making a lot of consumption decisions at a variety of designs z. Clearly, we can identify her demand function if (i) her preferences are constant over time and we observe a large panel or experiment,[8] or (ii) she belongs to a large homogeneous segment of consumers with identical preferences. This explains the form of Assumption

[7] The realizations of a Dirichlet process are, almost surely, discrete distributions. Although I assumed continuity to prove the existence of a unique demand system in Section 3.3, these realizations can approximate any continuous distribution. This discrepancy has no practical implications.

[8] In this case, when the number of dates T is large, we can have a segment m for each consumer i.

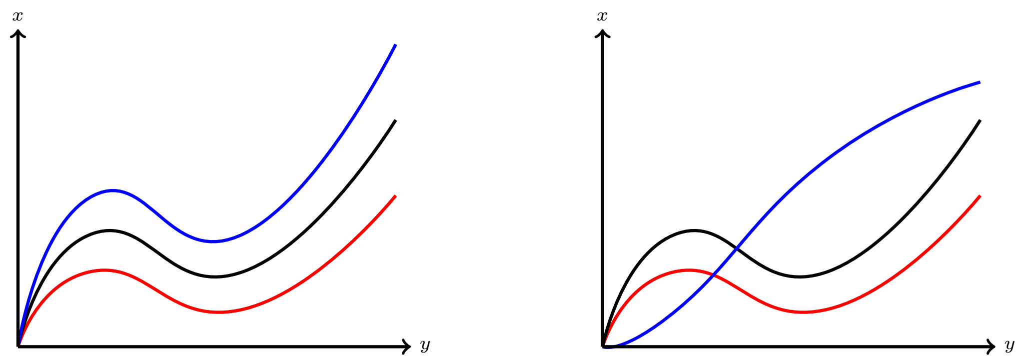

Figure 3.3. **Monotonicity in Heterogeneity.** Each figure displays the Engel curves for three consumers at a fixed price. Monotonicity is satisfied on the left. Monotonicity is violated on the right because the curves cross.

3.1 (as it allows for either interpretation). Later, we apply the segmented approach to scanner data in the application to the consumption of alcohol in Section 3.6.

With panel data, one no longer requires the assumption that demand is monotonic with respect to unobserved heterogeneity in order to achieve identification (see Figure 3.3, and the role of this assumption in Brown and Matzkin, 1995, Matzkin, 2003, and Hausman and Newey, 2016b).

3.5.1 Within Model Identification

In the models introduced in Sections 3.2 and 3.3, and for any π such that demand is invertible, we can derive the inverse demand function, whose second component coincides with the MRS which can be integrated to obtain a unique preference ordering. Indeed, by construction, the integrability conditions (needed to recover a *unique* well-behaved preference ordering) are satisfied, implying that preferences are recoverable (see Samuelson, 1948, for a seminal discussion of integrability in the case of two goods, and Samuelson, 1950, Hurwicz and Uzawa, 1971, and Hosoya, 2016, for general approaches). However, the possibility to recover preferences from a consumer's demand function does not imply that the distribution of taste uncertainty π is identified: Two distinct taste distributions could produce the same MRS.

For identification, I only consider the information contained in the demand function $X(\cdot;\pi)$ on the set $\mathcal{Z}$ of designs z for which the components of the demand function are strictly positive. This restriction disregards some information that may be available in the first or third regimes of (3.2.15). In most datasets, when a component of the demand function equals zero, the price p is not observed.

Stochastic Absolute Risk Aversion

In the SARA model, the identification condition is:

$$\left\{\frac{\mathbb{E}_{\pi}[A_1\exp(-A'x)]}{\mathbb{E}_{\pi}[A_2\exp(-A'x)]} = \frac{\mathbb{E}_{\pi'}[A_1\exp(-A'x)]}{\mathbb{E}_{\pi'}[A_2\exp(-A'x)]},\ \forall x\in R\right\} \ \Rightarrow\ \pi=\pi'. \tag{3.5.1}$$

In the degenerate case in which A is deterministic and equal to (a_1, a_2), the MRS reduces to a_1/a_2. Thus, in this special case, the two-dimensional parameter $a = (a_1, a_2)$ is identified up to a positive factor. This reasoning leads us to a question: Does this lack of identification also exist in an extended setting?

First, notice that, the utility function $U(x;\pi)$ is equal to the moment generating function for π with a negative sign: $\Phi(x;\pi) = -U(x;\pi)$. Because this moment generating function characterizes π when the stochastic parameter A is non-negative (see Theorem 1a in Chapter 13 on Tauberian Theorems in Feller, 1968), it is equivalent to consider the identification of either π, or $\Phi(x;\pi)$.[9] As mentioned, we can always integrate the MRS to recover a unique preference ordering. That is, we can recover $U(x;\pi)$ up to a monotonic transformation. We still need to discern the conditions on π under which we can recover $\Phi(x;\pi)$. Indeed, moment generating functions have properties that are not necessarily preserved under monotonic transformations.

We obtain the following result:

Proposition 3.5. If preferences are SARA, then $\Phi(x;\pi)$ and $\Phi(x;\pi)^\nu$ lead to the same preference ordering, for all positive scalars $\nu > 0$.

Proof. Let $U(x;\pi) = -\Phi(x;\pi)$ and $\tilde{U}(x;\pi) = -\Phi(x;\pi)^\nu$ denote the utility functions associated with $\Phi(x;\pi)$ and $\Phi(x;\pi)^\nu$, respectively. Then, by definition, we must have:

$$\tilde{U}(x;\pi) = -\Phi(x;\pi)^\nu = -(-U(x;\pi))^\nu = \phi_\nu(U(x;\pi)), \tag{3.5.2}$$

where $\phi_\nu(u) = -(-u)^\nu$ is strictly increasing for $u < 0$. Since $\tilde{U}(x;\pi)$ is a monotonic transformation of $U(x;\pi)$, these utility functions yield the same preference ordering. $\square$

This means that we can, at most, identify the class of moment generating functions $\mathscr{C}(\Phi) = \{\Phi^\nu : \nu > 0\}$. Note that, for any moment generating function Φ, the transformed function Φ^ν is also a moment generating function.

Let us now consider identification when A_1 and A_2 are independent:

Proposition 3.6. Let Φ_j denote the marginal moment generating function for A_j, for $j = 1, 2$. If preferences are SARA, and A_1 and A_2 are independent, then (Φ_1, Φ_2) and (Φ_1^*, Φ_2^*) lead to the same preference ordering if, and only if, for some $\nu > 0$, we have:

$$\Phi_1^* = \Phi_1^\nu \;\text{ and }\; \Phi_2^* = \Phi_2^\nu.$$

Proof. The identification criterion becomes:

$$\left(\frac{\partial \Phi_1(x_1)}{\partial x_1}\Phi_2(x_2)\right)\left(\Phi_1(x_1)\frac{\partial \Phi_2(x_2)}{\partial x_2}\right)^{-1} = \left(\frac{\partial \Phi_1^*(x_1)}{\partial x_1}\Phi_2^*(x_2)\right)\left(\Phi_1^*(x_1)\frac{\partial \Phi_2^*(x_2)}{\partial x_2}\right)^{-1},$$

[9] The existence of the moment generating function does not imply the existence of all power moments and, even if all power moments exist, they do not necessarily characterize the distribution. A known example is the log-normal distribution used in the application (Heyde, 1963).

for all $x \in R$. This criterion can, then, be written as:

$$\frac{\partial \log \Phi_1(x_1)}{\partial x_1}\left(\frac{\partial \log \Phi_1^*(x_1)}{\partial x_1}\right)^{-1} = \frac{\partial \log \Phi_2(x_2)}{\partial x_2}\left(\frac{\partial \log \Phi_2^*(x_2)}{\partial x_2}\right)^{-1},$$

for all $x \in R$. Thus, we deduce that, if these distributions yield the same MRS, then:

$$\frac{\partial \log \Phi_j^*(x_j)}{\partial x_j} = \nu \frac{\partial \log \Phi_j(x_j)}{\partial x_j},$$

for some $\nu > 0$, at every $x_j \geq 0$, for both $j = 1, 2$. Because the log-transform of the moment generating function at zero equals zero, by integrating this equation, we get:

$$\log \Phi_j^*(x_j) = \nu \log \Phi_j(x_j), \tag{3.5.3}$$

at every $x_j \geq 0$, for both $j = 1, 2$. Equivalently, $\Phi_1^* = \Phi_1^\nu$ and $\Phi_2^* = \Phi_2^\nu$. □

Proposition 3.6 implies that $\mathscr{C}(\Phi)$ is identified under the independence of A_1 and A_2. Indeed, we can recover the consumer's preference ordering using traditional methods, and use the fact that all admissible preference orderings map to a unique class $\mathscr{C}(\Phi)$.

Of course, independence is a strong restriction. In the SARA model, it is equivalent to the additive separability of the utility function.[10] To see this result, notice that, under independence, we must have:

$$U(x;\pi) = -\mathbb{E}_\pi\big[\exp(-A_1x_1)\big] \cdot \mathbb{E}_\pi\big[\exp(-A_2x_2)\big]. \tag{3.5.4}$$

Since utility functions are unique up to strictly increasing transformations, this utility function is equivalent to:

$$\tilde{U}(x;\pi) \equiv -\log(-U(x;\pi)) = -\log \mathbb{E}_\pi\big[\exp(-A_1x_1)\big] - \log \mathbb{E}_\pi\big[\exp(-A_2x_2)\big], \tag{3.5.5}$$

which is an additively separable utility function. In Appendix 3.E, I prove a generalization of Proposition 3.6, where stochastic taste parameters have a common component.

Stochastic Safety-First

In the SSF model, the identification condition is:

$$\left\{\frac{\mathbb{E}_\pi\big[A_1 \big| x_1 + A_1x_2 - A_2 < 0\big]}{\mathbb{E}_{\pi'}\big[A_1 \big| x_1 + A_1x_2 - A_2 < 0\big]} = 1,\ \forall x \in R\right\} \ \Rightarrow\ \pi = \pi'. \tag{3.5.6}$$

Let us now consider the validity of this condition under an independence assumption. Note, in the SSF model, independence is no longer equivalent to additive separability.

[10] In the case of two goods, additive separability is stronger than separability.

Proposition 3.7. If preferences are SSF, A_1 and A_2 are independent, and the marginal distribution of A_2 is continuous, then $\mathbb{E}[A_1]$ is identified, and the marginal distribution of A_2 is identified up to some positive power transformation of its survival function.

Proof. In the SSF model, the MRS is identified, and it satisfies:

$$\mathbb{E}_\pi[A_1 S(x_1 + A_1 x_2)] = \mathrm{MRS}(x;\pi)\mathbb{E}_\pi[S(x_1 + A_1 x_2)], \tag{3.5.7}$$

where $S(\cdot)$ denotes the survival function of A_2.

(i) The expectation $\mathbb{E}_\pi[A_1]$ is identified because $\mathrm{MRS}(x_1, 0;\pi) = \mathbb{E}_\pi[A_1]$.

(ii) By differentiating (3.5.7) with respect to x_2, we get:

$$\begin{aligned}\mathbb{E}_\pi[A_1^2 S'(x_1 + A_1 x_2)] = \mathrm{MRS}(x;\pi)\mathbb{E}_\pi[A_1 S'(x_1 + A_1 x_2)]\\ +\frac{\partial \mathrm{MRS}}{\partial x_2}(x;\pi)\mathbb{E}_\pi[S(x_1 + A_1 x_2)].\end{aligned}$$

When $x_2 = 0$, this equation becomes:

$$S'(x_1)\mathbb{E}_\pi[A_1^2] = \mathrm{MRS}(x_1, 0;\pi)S'(x_1)\mathbb{E}_\pi[A_1] + \frac{\partial \mathrm{MRS}}{\partial x_2}(x_1, 0;\pi)S(x_1).$$

By rearranging, we get:

$$\frac{\partial \mathrm{MRS}}{\partial x_2}(x_1, 0;\pi) = \frac{S'(x_1)}{S(x_1)}\left(\mathbb{E}_\pi[A_1^2] - \mathrm{MRS}(x_1, 0;\pi)\mathbb{E}_\pi[A_1]\right) = \frac{S'(x_1)}{S(x_1)}V(A_1).$$

Because the partial derivative of the MRS with respect to x_2 is identified, the hazard function $\lambda(x_1) = -S'(x_1)/S(x_1)$ of the distribution of A_2 is identified up to a positive factor. Since $S(x_1) = \exp\{-\Lambda(x_1)\}$, where $\Lambda(x_1) = \int_0^{x_1}\lambda(t)dt$ is the cumulative hazard function of the distribution of A_2, we can identify $S(\cdot)$ up to a positive power transformation.

□

Proposition 3.7 provides no information on the identifiability of the distribution of A_1 beyond its first moment. It seems difficult to obtain a general identification result, but insights into the identification problem can be obtained by considering the two primary families of distributions that are invariant to positive power transformations, that are, the exponential family and the Pareto family, respectively.

(i) **Exponential family:** Suppose that the marginal distribution of A_2 belongs to the exponential family, and that we have identified its survival function up to a positive power transformation such that $S(x) = \exp\{-cx\}$, for some unknown $c > 0$. The MRS in (3.5.7) becomes:

$$\mathrm{MRS}(x;\pi) = \frac{\mathbb{E}_\pi[A_1 \exp\{-cx_2 A_1\}]}{\mathbb{E}_\pi[\exp\{-cx_2 A_1\}]} \equiv G_0(x_2). \tag{3.5.8}$$

This expression does not depend on x_1. Now, let $\Psi(u) = \mathbb{E}_\pi[\exp\{-uA_1\}]$ denote the Laplace transform of A_1. Under this notation, the equality in (3.5.8) implies:

$$G_0(x_2) = \frac{d\log\Psi}{du}(cx_2).$$

Or, equivalently, $G_0(u/c) = d\log\Psi(u)/du$. By integrating, we obtain:

$$\log\Psi(u) = c\big[H(u/c) - H(0)\big],$$

where $H(\cdot)$ is a primitive of the MRS. Therefore:

Corollary 3.1. Under the conditions of Proposition 3.7, if the marginal distribution of A_2 belongs to the exponential family, the following results hold:

(a) The power transform c is not identified.

(b) The distribution of A_1 is identified under an identification restriction on c.

It is concluded that, under the conditions of Corollary 3.1, the distributions of A_1 and A_2 are non-parametrically identified up to a single scalar parameter $c > 0$.

(ii) **Pareto family:** Let us now examine whether a similar result can be obtained for the Pareto family, in which $S(x) = x^{-\alpha}$, for some $\alpha > 0$. The parameter α characterizes the fat tails of the distribution of A_2 and the power transformation on the MRS. This survival function produces:

$$\mathrm{MRS}(x;\pi) = \frac{\mathbb{E}_\pi[A_1(x_1 + A_1x_2)^{-\alpha}]}{\mathbb{E}_\pi[(x_1 + A_1x_2)^{-\alpha}]} = \frac{\mathbb{E}_\pi[A_1(x_0 + A_1)^{-\alpha}]}{\mathbb{E}_\pi[(x_0 + A_1)^{-\alpha}]},$$

where $x_0 \equiv x_1/x_2$ denotes a ratio of quantities. Equivalently, we get:

$$\mathrm{MRS}(x;\pi) = \frac{\mathbb{E}_\pi[(x_0 + A_1)^{-\alpha+1}]}{\mathbb{E}_\pi[(x_0 + A_1)^{-\alpha}]} - x_0 \equiv G_0(x_0), \tag{3.5.9}$$

which only depends on the ratio x_0. Therefore, we have constructed homothetic preferences. By equation (3.5.9):

$$e(x) \equiv \frac{d}{dx}\log\mathbb{E}_\pi[(x + A_1)^{-\alpha+1}],$$

is identified up to a multiplicative constant. Therefore, by integration, $\mathbb{E}_\pi[(x+A_1)^{-\alpha+1}]$ is identified up to α and a multiplicative constant κ. However, as x tends to infinity, this expression is equivalent to $\kappa x^{-\alpha+1}\exp E(x)$, where $E(\cdot)$ is a primitive of $e(\cdot)$. This tail behaviour provides both the identification of α and κ. This analysis is summarized by the following result:

Corollary 3.2. Under the conditions of Proposition 3.7, if the marginal distribution of A_2 belongs to the Pareto family, the distributions of A_1 and A_2 are non-parametrically identified.

3.5.2 Between Model Identification

Once the identification of the consumer's taste distribution π within each model is solved, we still need to consider the identification between the models. This analysis is needed to test whether preferences are consistent with SARA, or SSF, or both. It is important to know whether these two classes of preferences are nested or non-nested. If they are non-nested, we need to characterize their intersection and define a general class encompassing both types of preferences.

To illustrate, suppose that the consumer has SSF preferences:

$$U(x;\pi) = \mathbb{E}_\pi\big[\min\{x_1 + A_1 x_2,\, A_2\}\big], \tag{3.5.10}$$

where (i) A_1 and A_2 are independent, (ii) A_1 has distribution π_2, and (iii) A_2 follows an exponential distribution (with unit intensity). Under this specification, we obtain:

$$U(x;\pi) = 1 - \mathbb{E}_\pi\big[\exp(-x_1 - A_1 x_2)\big]. \tag{3.5.11}$$

To clarify this result, observe that, by conditioning on A_1, we are left with the expectation of the minimum of a set containing a constant and a random variable with an exponential distribution. This utility function is a strictly increasing transformation of a SARA utility function:

$$\tilde{U}(x;\pi) = -\mathbb{E}_\pi\big[\exp(-B'x)\big], \tag{3.5.12}$$

where (i) B_1 follows a point mass at 1, and (ii) B_2 has distribution π_2. Consequently, these utility functions, one SARA, and the other SSF, induce the same preference ordering over the consumption set.

3.5.3 Discussion

The possible lack of identification of each consumer's taste distribution π_m has to be taken into account in the economic interpretation of the results. However, it has to be noted that it does not create difficulties for structural inference, where the (scalar or functional) parameters of interest are the parameters characterizing the MRS, rather than the parameters characterizing the utility function.

The lack of identification is due to the special structure of the cone of increasing and concave functions defined on R, and of the extremal elements of this cone. For finite increasing concave functions defined on $\mathbb{R}_+$, it is well-known that the extremal functions are of the type:

$$h_0(x) = \min\{\alpha_1 x + \beta_1,\, \alpha_2 x + \beta_2\}, \tag{3.5.13}$$

in which $(\alpha_j, \beta_j) \in \bar{R}$, for $j = 1, 2$ (see Blaschke and Pick, 1916), and that any finite positive increasing concave function can be written as:

$$b + \mathbb{E}_\pi\big[\min(A, x)\big], \tag{3.5.14}$$

where b is a positive scalar and π is the distribution of A. Such functions are characterized by b and π. The set of extremal functions in (3.5.14) is a minimal set of extremal points generating the cone.

Such a property no longer holds for finite positive increasing concave functions defined on $\bar{R}$. Johansen (1974) has described a large set of extremal points of the type:

$$h_1(x) = \min\big\{\alpha_1 x + \beta_1, \ldots, \alpha_n x + \beta_n\big\}, \tag{3.5.15}$$

for which $h_1(\cdot)$ induces a covering with vertices of order 3 (see page 62 in Johansen, 1974), and has shown that this set is dense in the cone of finite continuous convex functions defined on a convex set in R (see Theorem 2 in Johansen, 1974). A minimal set of extremal points generating this cone does not exist. This argument explains why Sections 3.2 and 3.3 consider specific convex subsets generated by parametric functions.

While we restrict our attention to SARA and SSF preferences (because the stochastic taste parameters have clear interpretations in these models), other convex hulls could have been considered. For example:

(i) The convex hull generated by the union of the SARA and SSF models—that is, the smallest structural model containing both of the models in Sections 3.2 and 3.3.

(ii) The convex hull generated by a basis of the form:

$$U(x; a, \nu) = \frac{a_1}{\nu_1} x_1^{\nu_1} + \frac{a_2}{\nu_2} x_2^{\nu_2}, \tag{3.5.16}$$

for every $x \in \bar{R}$ in which $a \in R$ and $\nu \in (0,1)^2$. This basis corresponds to a first-order expansion of a utility function (see Johansen, 1969), and contains a Stone-Geary utility function as a limiting case. Indeed, as ν approaches zero, we obtain: $U(x; a) = a_1 \log x_1 + a_2 \log x_2$. However, the convex hull generated by this basis is not flexible enough because it only contains weighted combinations of $x_1^{\nu_1}$ and $x_2^{\nu_2}$. Similarly, the convex hull generated by a Stone-Geary basis only contains Stone-Geary utility functions, where the weights are the means of the taste parameters:

$$U(x; \pi) = \mathbb{E}_\pi\big[A_1 \log x_1 + A_2 \log x_2\big] = \mathbb{E}_\pi\big[A_1\big] \log x_1 + \mathbb{E}_\pi\big[A_2\big] \log x_2. \tag{3.5.17}$$

The Stone-Geary basis $U(x; a)$ above can be adjusted to define another parametric basis. In particular, consider the transformation $\varphi(x) = -\exp(-x)$ to the Stone-Geary utility function. This transformation yields:

$$\tilde{U}(x; a) \equiv \varphi(U(x; a)) = -\frac{1}{x_1^{a_1} x_2^{a_2}}. \tag{3.5.18}$$

This utility function forms a well-behaved basis because it is strictly increasing with a negative semi-definite Hessian. While $U(x;a)$ and $\tilde{U}(x;a)$ represent the same preference ordering, they will generate different families due to the strict concavity of $\varphi(\cdot)$. To illustrate, suppose that the stochastic parameters, A_1 and A_2, are independently distributed with respect to uniform distributions on $[0,1]$. This specification produces:

$$U(x;\pi) = \frac{1}{2}\log x_1 + \frac{1}{2}\log x_2 \ \text{ and } \ \tilde{U}(x;\pi) = -\left(\frac{x_1-1}{x_1\log x_1}\right)\left(\frac{x_2-1}{x_2\log x_2}\right). \tag{3.5.19}$$

While $U(x;\pi)$ is a Stone-Geary utility function, $\tilde{U}(x;\pi)$ is a complicated non-linear function of x_1 and x_2. Consequently, an uninteresting basis has been transformed into an interesting one. This procedure can be completed for any increasing, concave, and twice-differentiable transformation $\varphi(\cdot)$.

3.6 An Illustration

This section shows how to use the SARA and SSF models in a non-parametric framework. First, I specify the statistical model by introducing an assumption on the observations, and then we discuss statistical inference. The methodology is illustrated in an application to alcohol consumption using scanner data concerning individual purchase histories.

3.6.1 Assumptions on Observations

The behavioural models introduced in the previous sections can be completed with an assumption on the available observations. I consider panel data, indexed by the consumer i and date t. After a preliminary treatment of the purchase histories, we obtain a large number n of consumers and a fixed number T of observed dates. In the preliminary treatment, the goods are aggregated into two groups using a common quantity unit and the dated purchases are aggregated by month (see Section 3.6.3). Recall that, under Assumption 3.1, we have M distinct segments of homogeneous consumers.

I introduce the following assumption on the observations:

Assumption 3.2 (Observations)**.**

(i) We jointly observe (x_{it}, z_{it}), for all $i = 1,\ldots,n$ and $t = 1,\ldots,T$, when $x_{it} > 0$.

(ii) The individual histories $(x_{it}, z_{it})_{t=1}^{T}$ are independent given all π_m, $m = 1,\ldots,M$.

(iii) Designs (z_{it}) are exogenous (independent of taste distributions π_m).

Assumption 3.2 describes the structure of the observations. It implies that we can imagine taste parameters (π_m) being independently drawn from a Dirichlet process F, designs (z_{it}) being independently drawn from some distribution, and consumption x_{it} satisfying $x_{it} = X(z_{it};\pi_{m_i})$, where m_i is the group of consumer i. Many papers assume that consumption x_{it} is positive (see Section IV.A in Blundell, Horowitz, and Parey, 2017, for this assumption in an application to gasoline demand, as well as Assumption 2.4 in Chapter 2, for this assumption in an application to the consumption of alcohol); the SARA and SSF models allow for corner solutions. However, in many datasets (including the dataset used in the application in Section 3.6), there is a problem of partial observability. Let $\tilde{y}$ denote the expenditure (prior to normalization), and let $\tilde{p}_j$ denote the price of good j (prior to normalization). Usually, we only observe the price $\tilde{p}_j$ of a good j when the consumer buys a positive quantity of good j. Then, we only observe (normalized) expenditure y when the consumer buys a positive quantity of good 2, and we only observe the (normalized) price p when the consumer buys a positive quantity of both goods (see Crawford and Polisson, 2016, for an approach to revealed preference that deals with this partial observability problem). This problem explains the specific form of Assumption 3.2(i).

For deriving the asymptotic properties of estimators, it is also necessary to specify the type of asymptotics to be considered:

Assumption 3.3. Let n_m denote the size of the m^{th} homogeneous group.

(i) $n_m T \to \infty$, as $n \to \infty$, for all $m = 1, \ldots, M$.

(ii) $n_m T \sim \lambda_m n$, for some $\lambda_m \in (\lambda_\ell, \lambda_h)$, where $0 < \lambda_\ell < \lambda_h < 1$, for $m = 1, \ldots, M$.

(iii) $M \to \infty$, as $n \to \infty$.

Assumptions 3.3(i) and 3.3(ii) ensure that there are enough observations to non-parametrically estimate the demand function associated with the functional parameter π_m on a sufficiently large subset $\mathcal{Z}_m$ of designs z. Assumption 3.3(iii) guarantees enough filtered parameters $\hat{\pi}_m$ to estimate the underlying Dirichlet process F. In some special circumstances, T is large, and Assumption 3.3 can be used with $m = i$ and $M = n$—that is, a single consumer per group. Otherwise, grouping of homogeneous consumers is needed to identify the demand functions on sufficiently large subsets $\mathcal{Z}_m$.

3.6.2 Estimation Method

The Dirichlet process is common in Bayesian estimation (see, for instance, Ferguson, 1974, and Li et al., 2019). This process is useful because it is flexible and, if observations are independently and identically drawn from an unknown distribution, the posterior distribution of this distribution has a closed-form expression. That being said, the current framework is much more complicated for two reasons:

(i) The observed consumption choices (X_{ijt}) are not identically distributed because consumers make decisions at *different expenditures and prices.*

(ii) It is difficult to derive a closed-form expression for the demand, as a function of the expenditure, the price, and the functional parameter characterizing taste uncertainty π. It is, therefore, difficult to derive a closed-form expression for the distribution of X_{ijt} conditional on Z_{it}.

These features of the model explain why estimation requires specific numerical algorithms. These specific algorithms have to be able to deal with the non-linear and high-dimensional features of the models. In the Bayesian framework, the Dirichlet process is fixed. In the hyperparametric framework, it is parameterized by a vector θ. These parameters have to be estimated and the functional parameters (π_m) have to be filtered. These estimation approaches are described below.

Bayesian Framework

In a pure Bayesian framework, a Dirichlet process is fixed by selecting a mean distribution μ and a scaling parameter c (see Appendix 3.A). This distribution defines the common prior for the functional taste parameters (π_m). After, the data are used to compute the posterior distribution for the functional parameters (π_m). Under Assumption 3.2, the posterior distribution can be computed separately for each homogeneous group of consumers:

$$\ell(\pi_m | x_{it}, z_{it}, x_{it} > 0, i \in \Lambda_m, t = 1, \ldots, T),$$

where Λ_m denotes the group of consumers with preferences characterized by the taste parameter π_m. This approach does not have to account for the potential identification problem discussed in Section 3.5. If a specific characteristic of π_m is weakly identified, its posterior distribution will be close to the prior distribution.

In the current framework, the observations (x_{it}, z_{it}), conditional on $x_{it} > 0$, must satisfy the deterministic first-order conditions implied by the model. These conditions have the following form:

$$\text{MRS}(x_{it}; \pi) = p_{it}, \tag{3.6.1}$$

for any observed pair (x_{it}, z_{it}). Equivalently:

$$\mathbb{E}_\pi \left[\frac{\partial U(x_{it}; A)}{\partial x_1} \right] = p_{it} \, \mathbb{E}_\pi \left[\frac{\partial U(x_{it}; A)}{\partial x_2} \right], \tag{3.6.2}$$

for any observed pair (x_{it}, z_{it}). These conditions are moment restrictions, called *MRS restrictions*. In the current big data framework, the number of MRS restrictions is very large, typically several hundred to a thousand. The posterior of π_m is simply the distribution of π_m given these deterministic restrictions on π_m. If the taste parameters, A_1 and A_2, are independent with marginal distributions, π_1 and π_2, respectively, then the MRS restrictions are bilinear in π_1 and π_2—specifically, these restrictions are linear in π_1 given π_2, and linear in π_2 given π_1. Later, this property is used to construct a numerically efficient optimization algorithm for filtering all the π_m (see Appendix 3.C).

Hyperparametric (or Empirical Bayesian) Framework

The hyperparametric (or empirical Bayesian) framework is a complicated non-linear state-space model with two layers of latent state variables. Such a framework can be characterized as follows:

(i) **Deep layer:** Functional parameters (π_m), drawn from F (parameterized by θ);

(ii) **Surface layer:** Demand functions $X(\cdot;\pi_m)$ deduced from π_m;

(iii) **Measurement equations:** Observed pairs (x_{it}, z_{it}), given $x_{it} > 0$.

We have partial observability of the demand function because the value of demand $X(z;\pi_m)$ is observed at finitely many designs z. Furthermore, unlike most state-space models, the state variables are infinite-dimensional.

Estimating the Hyperparameter

While it is difficult to derive analytically the distribution of X_{it} given Z_{it}, it is easy to simulate its distribution for a given value of θ (see Appendix 3.A for simulations from the Dirichlet distribution). Therefore, θ can be estimated by the method of simulated moments (MSM), or indirect inference (see McFadden, 1989, Pakes and Pollard, 1989, and Gouriéroux and Monfort, 1996). That is, θ is estimated by matching some sample and simulated moments of the pair (X_{it}, Z_{it}).

To illustrate, consider a pure panel such that $M = n$.[11] The steps are the following:

Step 1: Simulate $s = 1, \ldots, n$ draws from a Dirichlet process given the parameter θ. Each draw $\pi^s(\theta)$ is associated with an individual consumer i such that $s = i$.

Step 2: Compute simulated consumption $x^s_{it}(\theta)$ by solving the first-order condition in (3.2.12) with respect to x_1 and applying the transformation in (3.2.15) given $z_{it} = (y_{it}, p_{it})$ and $\pi^i(\theta)$.

Step 3: Construct a collection of K moments from the observed and simulated data:

$$m \equiv \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m_k(x_{it}, z_{it})\right]_k \quad \text{and} \quad m(\theta) \equiv \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m_k(x^s_{it}(\theta), z_{it})\right]_k .$$

Then, numerically solve the following problem:

$$\underset{\theta}{\operatorname{argmin}} \; \big|\big|m - m(\theta)\big|\big|, \tag{3.6.3}$$

in which $||\cdot||$ is a Euclidean norm with the form $||m||^2 = m'\Omega m$, for some positive-definite $K \times K$ matrix Ω.

[11] When $M < n$, we simulate $n_m T$ observations for the m^{th} draw from the Dirichlet process.

Given the estimated hyperparameter $\hat{\theta}$, the taste distributions (π_m) must be filtered. This step is equivalent to applying the Bayesian approach with the estimated Dirichlet distribution as the prior distribution (see Appendix 3.C).

Under Assumptions 3.1 to 3.3, the estimator for θ is consistent and asymptotically normal, and it converges at a speed of $1/\sqrt{nT}$. The derivation of the asymptotic properties of the filtered functional parameter $\hat{\pi}_m$ is out of the scope of this paper and left for future research.

Filtering the Taste Distributions

Once the hyperparameter θ is estimated, we can filter π_m by using the following steps:

Step 1: Draw a taste distribution $\tilde{\pi}_m$ from the Dirichlet process given $\hat{\theta}$. Then, by construction, the taste distribution $\tilde{\pi}_m$ is a draw from the prior distribution.

Step 2: Discretize $\tilde{\pi}_m$ on a grid of values for the taste parameters, A_1 and A_2. Let $\bar{\pi}_m$ denote the result. The aim of this step is to put $\tilde{\pi}_m$ on a grid for optimization.

Step 3: Solve the minimization problem:

$$\min_{\pi} \ ||\pi - \bar{\pi}_m|| \ \text{ s.t. MRS restictions (3.6.1) and unit mass restrictions.}$$

Let $\hat{\pi}_m^*$ denote the solution. This solution approximates a drawing from the posterior.[12]

Step 4: Replicate these steps to obtain a sequence of solutions: $\hat{\pi}^*_{m,s}$, $s = 1, \dots, S$, where S is the number of replications. The filtered $\hat{\pi}_m$ is obtained by averaging over all simulations such that:

$$\hat{\pi}_m = \frac{1}{S} \sum_{s=1}^{S} \hat{\pi}^*_{m,s}$$

This procedure involves a high-dimensional argument π_m, and a very large number of MRS restrictions. Indeed, we need several hundred grid points for π_m, and, in the application, we have about one-thousand MRS restrictions, for each $m = 1, \dots, M$. If the taste parameters, A_1 and A_2, are independent, this procedure can be numerically simplified by using the fact that these restrictions are bilinear (see Section 3.6.2 and Appendix 3.C).

3.6.3 The Data

I use the Nielsen Homescan Consumer Panel (NHCP). Nielsen provides a sample of households with barcode scanners. Households are asked to scan all purchased goods on the date of each purchase. The

[12] This solution can be thought of as a projection of $\tilde{\pi}_m$ onto the space of discrete probabilities which satisfy the MRS restrictions in (3.6.1). Imposing these restrictions ex ante would often lead to no solution. This aspect of the problem is related to finite tests of rationality (see Afriat, 1967).

Table 3.1. Mean m, standard deviation σ, the ratio σ/m, and quantiles for expenditure $\tilde{y}$, prices $\tilde{p}_j$, normalized expenditure y, and normalized price p. Normalized expenditures y and prices p are conditional on being defined.

				Quantiles					
Var.	m	σ	σ/m	0%	25%	50%	75%	100%	N
$\tilde{y}$	52.29	70.62	1.35	0.00	13.28	28.17	62.97	2,767.76	63,972
$\tilde{p}_1$	70.36	44.19	0.62	0.05	46.41	62.55	83.41	2,893.65	33,077
$\tilde{p}_2$	61.33	326.93	5.33	0.03	29.29	48.56	75.14	39,900.85	45,518
y	1.60	3.48	2.17	0.00	0.27	0.74	1.83	228.57	45,518
p	2.20	4.06	1.54	0.00	0.89	1.37	2.29	139.76	14,659

prices are entered by the households or linked to retailer data by The Nielsen Company. The households that agree to participate are compensated through benefits and lotteries.

I focus on the consumption of alcoholic drinks (see Manning et al., 1995, for an application to alcohol consumption in economics). I classify drinks by type. Good 1 contains beers and ciders.[13] Good 2 contains wines and liquors. I disregard all non-alcoholic beers, ciders, and wines. We are left with 30,635 beers and ciders, and 108,439 wines and liquors, for a total of 139,074 drinks. I convert all measurement units to litres of alcohol by first converting all units to litres and then multiplying by the standard alcohol by volume (ABV) in each subgroup—specifically, 4.5% for beer and cider, 11.6% for wine, and 37% for liquor. For example, if a household buys two packs of six bottles of beer and each bottle contains 355 millilitres of beer, then the household buys 4.26 litres of this beer, or $4.26 \times 0.045 = 0.231$ litres of alcohol. I use the standard ABV in each subgroup as a result of data limitations. The sample only contains purchases made at stores, not purchases made at bars, or restaurants.

Measuring quantities in litres of alcohol has at least three advantages: (i) it can account for a quality effect, (ii) it is appropriate for analyzing most relevant structural objects (e.g. the effect of a change in taxation on alcohol consumption), and (iii) it yields continuous quantities, permitting the application of standard tools in consumer theory (which could not be used if quantities were measured in, for example, bottles), and avoiding some common identification issues in the literature.

I restrict our sample to purchases made from August to November in 2016. This relatively short window is used to diminish the impact of changing tastes and product availability, and to avoid most federal holidays in the United States that are often associated with alcohol consumption such as Independence Day, Christmas Day, and New Year's Eve. The sample contains 28,036 households. Some additional details of this restricted sample are placed in Appendix 3.D.

The dated purchases are aggregated by month. For each household and month, the prices are constructed by dividing the total expenditure for each aggregate good (after accounting for the value of coupons) by the amount of alcohol of that aggregate good purchased by the household, when this

[13] The NHCP classifies ciders as wine, by default. I reclassify these beverages using product descriptions because most ciders have a low alcohol by volume (ABV).

Table 3.2. Proportions of observations by type.

	$x_2 = 0$	$x_2 > 0$
$x_1 = 0$	0.4298	0.2751
$x_1 > 0$	0.1642	0.1307

amount is strictly positive. Then, I normalize by the price of good 2. This procedure yields four monthly observations per household for a total of 112,144. A total of 63,936 observations have positive consumption such that $x_{it} > 0$. Table 3.1 gives summary statistics conditional on $x_{it} \neq 0$. The prices (p_{it}) are conditional on being well-defined (see the discussion of partial observability on pages 23 and 24). For the interpretation of the results, recall that $\tilde{y}$ is the expenditure (prior to normalization), and that $\tilde{p}_j$ is the price of good j (prior to normalization).

There are four regimes of observations: (i) zero expenditure on all goods, (ii) zero expenditure on good 1 and strictly positive expenditure on good 2, (iii) strictly positive expenditure on good 1 and zero expenditure on good 2, and (iv) strictly positive expenditure on all goods. Table 3.2 provides the proportion of observations in each regime, and shows a large proportion of observations with zero expenditure. Recall, under Assumption 3.2, designs z_{it} are drawn from a distribution. Therefore, we can interpret this result as a mass at zero in the marginal distribution of expenditure.

Figure 3.4 displays the sample distribution of expenditure $\tilde{y}_{it}$ by regime: the distribution of expenditure $\tilde{y}_{it}$ conditional on $x_{it} > 0$ is on the left; the sample distributions of expenditure $\tilde{y}_{it}$ for the two other regimes with positive expenditure are on the right. The shape of the sample distribution of expenditure $\tilde{y}_{it}$ does not appear to vary all that much with the regime. That being said, the sample distribution conditional on $x_{it} > 0$ has more probability attributed to higher expenditures.

Figure 3.5 compares the sample distributions of prices $\tilde{p}_j$ by regime: the sample distributions of $\tilde{p}_1$ are on the left; the sample distributions of $\tilde{p}_2$ are on the right. Although the sample distribution of $\tilde{p}_1$ differs from the sample distribution of $\tilde{p}_2$, these distributions do not seem to be affected by the regime.

Figure 3.6 displays the sample distributions of (normalized) designs $z_{it} = (y_{it}, p_{it})$ and the components of consumption x_{it} given $x_{it} > 0$. As expected, the components of consumption x_{it} are increasing in expenditure y_{it}. Furthermore, the first component of consumption x_{it} is more affected by changes in the price p_{it} than the second component.

Since I consider a rather short window of time, I follow the segmented population approach. I segment the population by state. Large states (e.g. California) are segmented again by county. Specifically, a county is given its own segment if it has more than 70 observations with positive consumption and it is in a state with more than 1,000 observations with positive consumption. We are left with a total of 65 segments, each corresponding to a state or county. The smallest segment is Wyoming, containing 15 observations with positive consumption; the largest state is Florida (after removing Broward, Hillsborough, Palm Beach, Pinellas, and Miami-Dade counties), containing 880 observations with positive

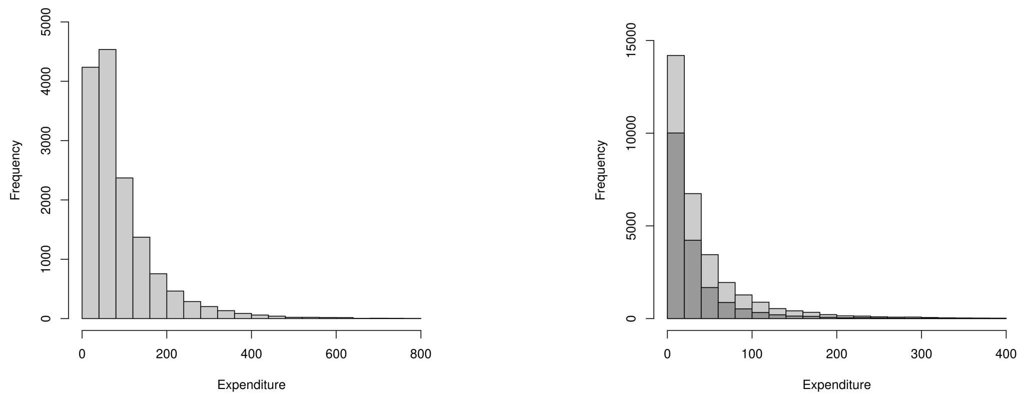

Figure 3.4. **Sample Distributions of Expenditure $\tilde{y}$ by Regime.** On the left, I illustrate the sample distribution conditional on $x_1 > 0$ and $x_2 > 0$; on the right, the light histogram illustrates the sample distribution conditional on $x_1 = 0$ and $x_2 > 0$, and the dark histogram illustrates the sample distribution conditional on $x_1 > 0$ and $x_2 = 0$.

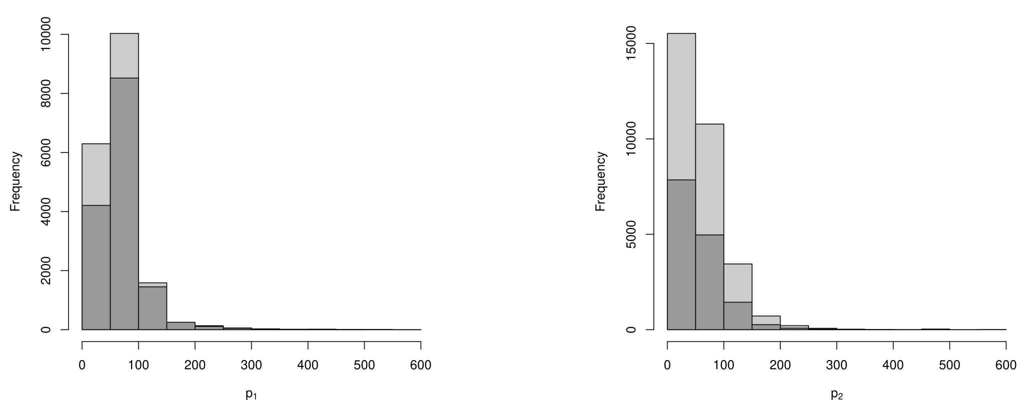

Figure 3.5. **Sample Distributions of Prices $\tilde{p}_j$ by Regime.** On the left, the light histogram illustrates the sample distribution of $\tilde{p}_1$ conditional on $x_1 > 0$ and $x_2 = 0$; on the right, the light histogram illustrates the sample distribution of $\tilde{p}_2$ conditional on $x_1 = 0$ and $x_2 > 0$; in each plot, the dark histogram illustrates the sample distribution conditional on $x_1 > 0$ and $x_2 > 0$.

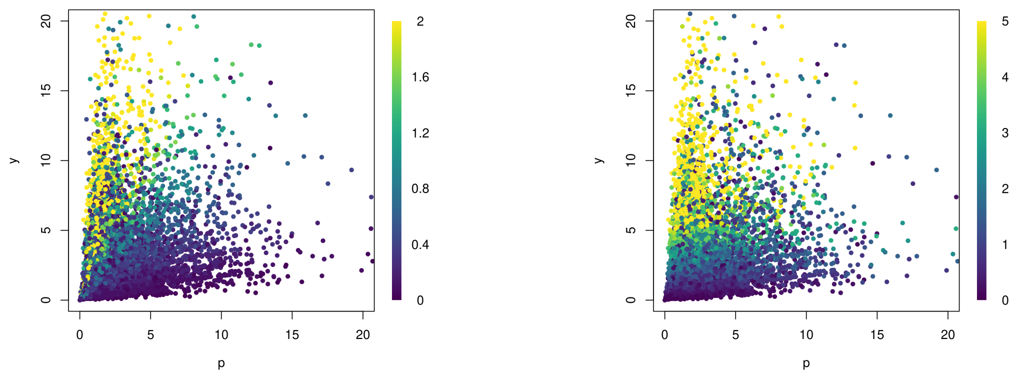

Figure 3.6. **Sample Distributions of Designs z_{it} and Consumption.** These figures are conditional on $x_{it} > 0$. On the left, colour describes the quantity of good 1; on the right, colour describes the quantity of good 2.

expenditure; the mean number of observations with positive consumption per segment is roughly 226.

Figure 3.7 displays the sample distributions of (normalized) designs $z_{it} = (y_{it}, p_{it})$ and the first component of consumption x_{it} given $x_{it} > 0$ in two of the larger segments: California (after removing Almeda, Los Angeles, Orange, Riverside, Sacramento, San Bernardino, and San Diego counties), and Florida (after removing Broward, Hillsborough, Palm Beach, Pinellas, and Miami-Dade counties).

Figure 3.8 displays the Nadaraya-Watson (kernel) estimates of the demand function for beer conditional on $x_{it} > 0$ in California and Florida over a subset of the domain of designs. Demand for beer in California is lower and less responsive to price changes than in Florida. Figure 3.9 displays Engel curves for good 1 in California and Florida given $p \equiv \tilde{p}_1/\tilde{p}_2 = 4$.[14] These Engel curves cross.

3.6.4 Estimation Results

As an illustration, I consider the SARA model in the hyperparametric framework. I assume that the taste parameters, A_1 and A_2, are independent. Under this assumption, the taste uncertainty is characterized by the marginal distributions, π_1 and π_2. The marginal distribution π_j of A_j is independently drawn from a Dirichlet process F_j, $j = 1, 2$. The mean of F_j is a log-normal distribution with parameters μ_j and σ_j, and the scale parameter of F_j is c_j. The utility function corresponding to this log-normal mean distribution, say $\bar{\pi}_j$, has a quasi closed-form expression. Indeed, under this distribution, we must have:

$$\log(A_j) = \mu_j + \sigma_j \varepsilon_j, \quad \forall j = 1, 2,$$

[14]This price is chosen to be in a sufficiently dense region of the sample distribution (see Figure 3.7).

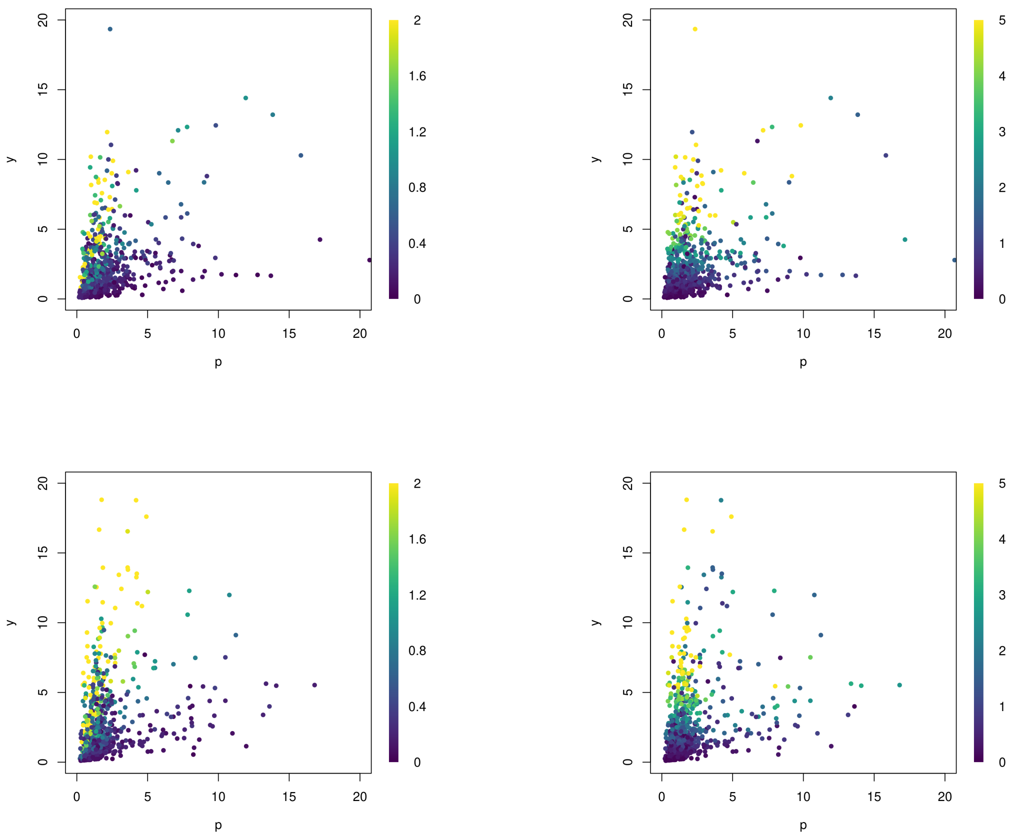

Figure 3.7. **Sample Distributions by State.** These figures are conditional on $x_{it} > 0$. California is shown on the top, and Florida is shown on the bottom. On the left, colour describes the quantity of good 1; on the right, colour describes the quantity of good 2.

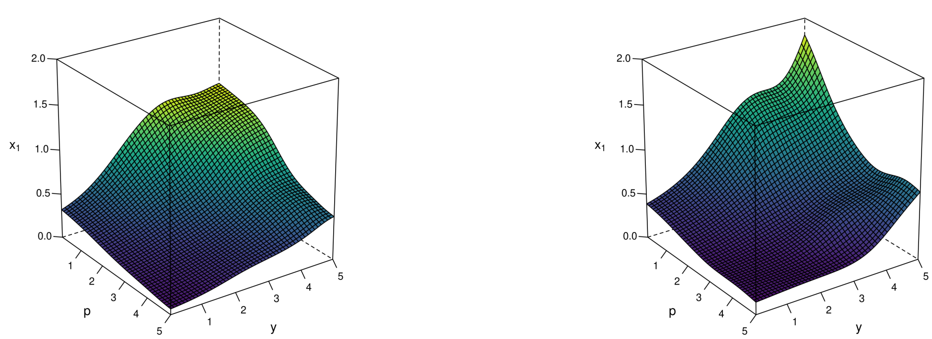

Figure 3.8. **Demand.** Nadaraya-Watson estimates of the demand function for good 1 conditional on $x_{it} > 0$ in California (left) and Florida (right).

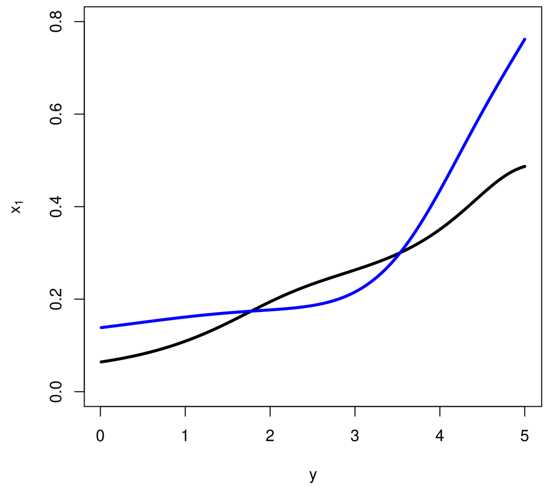

Figure 3.9. **Engel Curves.** Engel curves for good 1 in California (black) and Florida (blue). These curves violate monotonicity because they cross.

where ε_j is distributed with respect to a standard normal distribution. Then:

$$\mathbb{E}_{\bar{\pi}_j}[\exp(-A_j x_j)] = \mathbb{E}_{\bar{\pi}_j}[\exp(-\exp(\mu_j + \sigma_j \varepsilon_j) x_j)],$$
$$= \frac{1}{\sqrt{1 + w(x_j \exp(\mu_j)\sigma_j^2)}} \exp\left\{ -\frac{1}{2\sigma_j^2} w(x_j \exp(\mu_j)\sigma_j^2)^2 - \frac{1}{\sigma^2} w(x_j \exp(\mu_j)\sigma_j^2) \right\},$$

where $w(x)$ is the Lambert function, defined by the implicit equation:

$$w(x)\exp(w(x)) = x,$$

(see equation (1.3) in Asmussen et al., 2016). By drawing from the Dirichlet process, we will draw a stochastic utility function around the closed-form expression above. The hyperparameter θ has six components such that:

$$\theta = (\mu_1, \sigma_1, c_1, \mu_2, \sigma_2, c_2).$$

The Hyperparameter

As described in Section 3.6.2, the first step involves estimating the hyperparameter θ using the Method of Simulated Moments (MSM). The hyperparameter θ is calibrated by using the following (sample and simulated) moments computed for all of the 63,936 observations with positive consumption:

(i) marginal moments of (X_{it});

(ii) cross-moments of $(\log X_{it}, \log P_{it})$ and $(\log X_{it}, \log Y_{it})$;

(iii) cross-moments of (X_{it}, P_{it}), (X_{it}, Y_{it}), $(X_{it}, \log P_{it})$, and $(X_{it}, \log Y_{it})$.

The moments in (ii) are the moments used in the Almost Ideal Demand System (see Deaton and Muellbauer, 1980a); the moments in (iii) are introduced in order to capture risk effects by comparison with the moments in (ii). The optimum is found using a random search algorithm over a sufficiently big support.[15]

To apply MSM, it is necessary to compute simulated consumption $x_{it}^s(\theta)$ for every observation, at each step of the optimization algorithm. This procedure is computationally costly. Note that, the number of simulated observations with positive consumption is stochastic, and not necessarily equal to the number of observations with positive consumption in the sample. This aspect has no impact on the consistency of the MSM estimator.

The estimated hyperparameter $\hat{\theta}$ is:

$$\hat{\theta} = (0.7987, 3.5516, 45.0951, 0.1201, 3.6597, 3.5544). \tag{3.6.4}$$

Therefore, the median level of risk aversion for the mean of the Dirichlet process[16] for A_1 is $\exp(0.5495) \simeq 2.2226$, and the median level of risk aversion for the mean of the Dirichlet process for A_2 is $\exp(0.8738) \simeq 1.1276$. The fact that μ_1 is smaller than μ_2 is expected: Since quantities are measured in terms of volume of alcohol, this result is consistent with the faster overall intake of alcohol when consuming drinks with a higher ABV. Moreover, the distribution π_1 of A_1 is much more concentrated around its mean than the distribution π_2 of A_2, as the scaling parameter $c_1 = 45.0951$ for π_1 is much larger than the scaling parameter $c_2 = 3.5544$ for π_2.

I do not report any standard errors because they are automatically small from the large number of observations. Indeed, the standard significance test procedures (such as comparing a t-statistic to the critical value of a standard normal at the 5% significance level) are not relevant in this big data framework. The highest degree of uncertainty concerns the filtered functional parameters $(\hat{\pi}_m)$ since π_m is a high-dimensional parameter and the number of observations in each segment is much smaller.

The means of these Dirichlet processes are displayed in the left panel in Figure 3.10. The right panel displays the indifference curves associated with utility levels -0.1000, -0.0800, and -0.0680 for a draw from the Dirichlet process given $\hat{\theta}$.

Figure 3.11 displays the Q-Q plots for two draws (π_1^s, π_2^s), $s = 1, 2$, from the Dirichlet process given $\hat{\theta}$. In particular, I plot the quantiles of the realization of the distribution π_j of $\log(A_j)$ against the quantiles of the normal distribution given the estimated hyperparameters $(\hat{\mu}_j, \hat{\sigma}_j)$, for $j = 1, 2$. If these quantiles coincide exactly, they will lie on the 45-degree line. As expected, these Q-Q plots lie approximately around the 45-degree line. The draws (π_1^s), $s = 1, 2$, for π_1 are closer the 45-degree line and "more

[15] Random search is more efficient than grid search in hyperparameter optimization (Bergstra and Yoshua, 2012).

[16] This is not the absolute risk aversion of the utility function for the log-normal mean distribution which depends on the consumption level and has to be computed with a modified density.

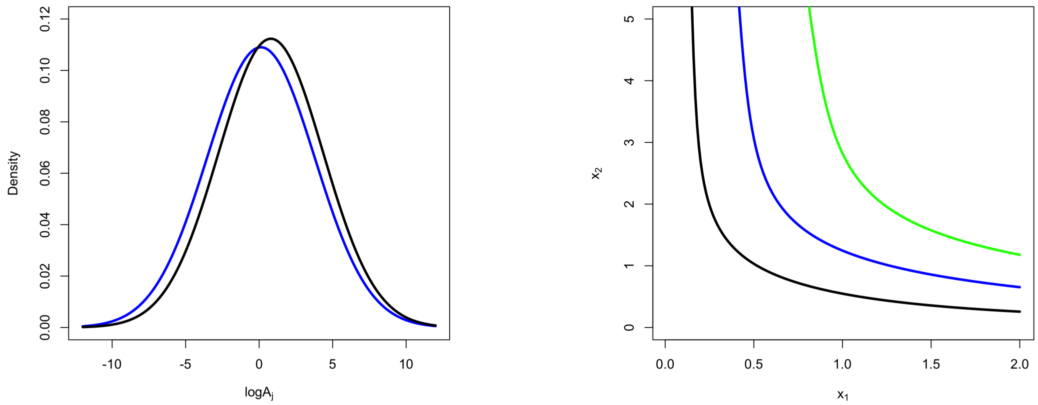

Figure 3.10. On the left, black shows the density of the mean of the Dirichlet process for A_1, and blue shows the density of the mean of the Dirichlet process for A_2. The x-axis is in log-scale. For scale: $\exp(-5) \simeq 0.0067$ and $\exp(5) \simeq 148.4131$. The figure on the right displays indifference curves associated with these distributions.

continuous” than the draws (π_2^s), $s = 1, 2$, for π_2 since $c_1 > c_2$.

Figure 3.11 illustrates how one might use the (estimated) hyperparameter for interpretation. Specifically, it is used to deduce the mean of the Dirichlet process, which is used as a benchmark for comparison with a drawn or filtered functional parameter π_m.

3.6.5 Taste Distributions

This section uses the filtering approach described in Section 3.6.2 to recover π_m. In the SARA model, the MRS restriction in (3.6.2) is:

$$\mathbb{E}_\pi[A_1 \exp(-A' x_{it})] = p_{it}\, \mathbb{E}_\pi[A_2 \exp(-A' x_{it})].$$

When A_1 and A_2 are independent, this expression becomes:

$$\begin{aligned} &\mathbb{E}_{\pi_1}[A_1 \exp(-A_1 x_{i1t})]\mathbb{E}_{\pi_2}[\exp(-A_2 x_{i2t})] \\ &= p_{it}\, \mathbb{E}_{\pi_1}[\exp(-A_1 x_{i1t})]\mathbb{E}_{\pi_2}[A_2 \exp(-A_2 x_{i2t})]. \end{aligned} \tag{3.6.5}$$

To filter π_m, these restrictions have to be imposed for every observation with positive consumption x_{it} associated with segment Λ_m. In California, there are 688 MRS restrictions, and, in Florida, there are 880. Appendix 3.C shows how to numerically solve the resulting optimization problem given the bilinearity of the MRS restrictions under independence.

The marginal taste distributions were filtered using a grid with 500 points between $\exp(-10)$ and $\exp(10)$, equally spaced on the log-scale. All draws from the estimated prior were simulated by the

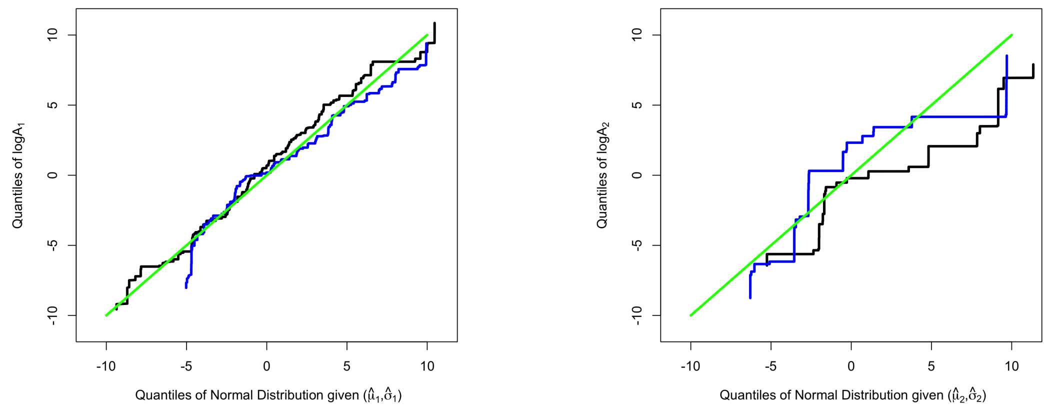

Figure 3.11. The Q-Q plots for two draws from the Dirichlet process given $\hat{\theta}$: On the left, the quantiles of $\log(A_1)$ are plotted against the quantiles of the normal distribution given $(\hat{\mu}_1, \hat{\sigma}_1)$; on the right, the quantiles of $\log(A_2)$ are plotted against the quantiles of the normal distribution given $(\hat{\mu}_2, \hat{\sigma}_2)$. In each figure, the green line is the 45-degree line.

stick-breaking method given $J = 1000$ breaks (see Appendix 3.A). Exactly $S = 100$ draws from the posterior were used to filter each distribution.

Figure 3.12 displays the Q-Q plots for the filtered taste parameters $\hat{\pi}_m$ for California and Florida. As in Figure 3.11, the (estimated) hyperparameter is used to construct a benchmark for comparison. As expected, the filtered taste parameters are rather different from this benchmark. Here, the role of the estimated prior distribution diminishes with the number of observations. In both states, the slope on the left is steeper than the 45-degree line, suggesting that the posterior mean distribution for A_1 is more "dispersed" than its estimated prior mean distribution. The convexity of these curves also suggests fatter tails.

For the structural interpretation of these plots, assume that (i) the preferences are SARA, (ii) the taste parameters are independent, (iii) the marginal distribution of A_1 is the same in both states, and (iv) the marginal distribution of A_2 "shifts" such that $\pi_2^*(A_2) = \pi_2(cA_2)$, where π_2 and π_2^* denote the marginal distributions of A_2 in these states. Under these assumptions:

$$U(x_1, x_2; \pi^*) = U(x_1, cx_2; \pi), \tag{3.6.6}$$

and solving the utility maximization problem in (3.2.13) yields:

$$X_1(z; \pi^*) = X_1(cz; \pi) \;\text{ and }\; X_2(z; \pi^*) = \left(\frac{1}{c}\right) X_2(cz; \pi). \tag{3.6.7}$$

Similarly, if there is a "shift" in the marginal distribution of A_1 and the marginal distribution of A_2 is

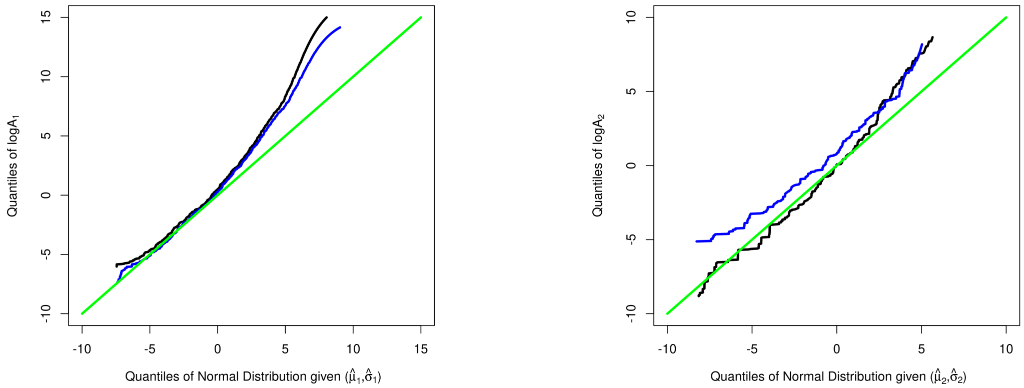

Figure 3.12. The Q-Q plots for the filtered taste distributions for California (black) and Florida (blue): On the left, the quantiles of $\log(A_1)$ are plotted against the quantiles of the normal distribution given $(\hat{\mu}_1, \hat{\sigma}_1)$; on the right, the quantiles of $\log(A_2)$ are plotted against the quantiles of the normal distribution given $(\hat{\mu}_2, \hat{\sigma}_2)$. In each figure, the green line is the 45-degree line.

the same in both states, we obtain:

$$X_1(z;\pi^*) = \left(\frac{1}{c}\right) X_1\left(y, \frac{p}{c};\pi\right) \quad \text{and} \quad X_2(z;\pi^*) = X_2\left(y, \frac{p}{c};\pi\right). \tag{3.6.8}$$

The relationships given in (3.6.7) and (3.6.8) suggest that there exists a complicated non-linear relationship between such demand functions. Therefore, we cannot immediately deduce from Figure 3.12 which state has a higher demand for beer. For a more formal analysis, the utility functions associated with each posterior mean taste distribution must be used to derive a posterior MRS, or a posterior demand function.

This analysis has to be completed with a discussion of accuracy. In this non-parametric framework, the posterior distributions of π_1 and π_2 are infinite-indimensional and cannot be represented. However, posterior distributions of any scalar transformation of π_1 and π_2 can be derived using simulation. In this respect, it is important to know which scalar objects are of interest. Typically, we are interested in the MRS evaluated at a specific bundle, say x_0, or counterfactual demand, corresponding to a particular design, say $z_0 = (y_0, p_0)$. Figure 3.13 displays the posterior distributions of the MRS, evaluated at two bundles, $(1, 1)$ and $(1, 2)$, for California and Florida. In both states, these distributions are approximately log-normal (with is consistent with Dobronyi and Gouriéroux, 2020), and the posterior for $\text{MRS}(1, 2; \pi)$ has a much longer tail than the posterior for $\text{MRS}(1, 1; \pi)$, implying that, the quantity of good 2 that must be given to the consumer in order to compensate her for one unit of good 2 (and keep her just as happy) is larger, on average, when she has more of good 2. This tail is longer in California.

The filtered taste distributions in Figure 3.12 are obtained by applying the algorithm in Appendix 3.C and forcing the density π_m to be non-negative at each iteration. The existence of negative "probabilities"

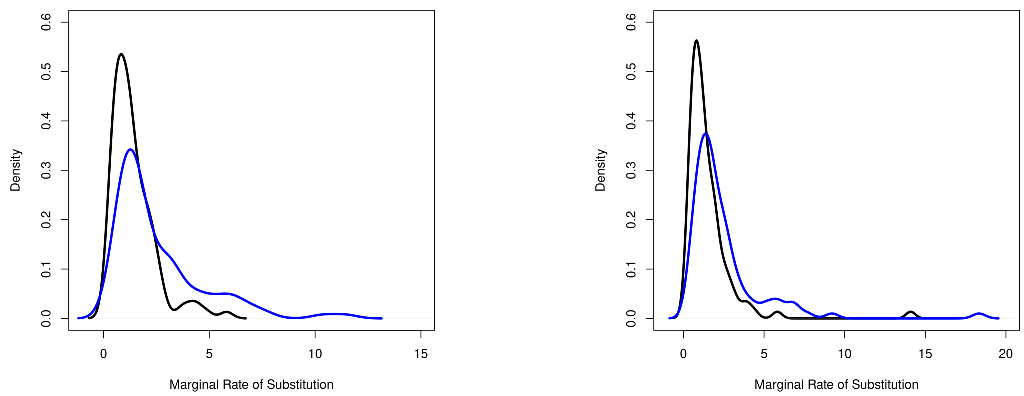

Figure 3.13. **Posterior Marginal Rate of Substitution.** The posterior distributions for $\mathrm{MRS}(1,1;\pi)$ (black) and $\mathrm{MRS}(1,2;\pi)$ (blue) for California (left) and Florida (blue).

can be a result of numerical uncertainty, the choice of grid, or misspecification. Specifically, it can arise if the consumer in segment Λ_m does not maximize her SARA/SSF utility function (or any utility function) subject to the linear budget constraint. By analyzing these negative probabilities, we can construct a measure of the deviation from rationality. To illustrate, let $\pi_k^+ = \max\{0, \pi_k\}$ and $\pi_k^- = \max\{0, -\pi_k\}$, respectively, denote the positive and negative components of the elementary probability π_k associated with the k^{th} grid point. The following ratio:

$$\mathrm{BR} = \frac{\sum_k \pi_k^-}{\sum_k (\pi_k^- + \pi_k^+)}, \tag{3.6.9}$$

is a measure of bounded rationality. This ratio ranges between 0 and 1. The closer this ratio is to 1, the less compatible the data are with the hundreds of MRS restrictions imposed by the chosen model. This ratio is related to a subset of the literature concerned with such measures. Existing measures include Afriat's Efficiency Index (Afriat, 1967; Varian, 1990), and the Money Pump Index (Echenique et al., 2011). In general, these indices are used to measure a single consumer's deviation from rationality by evaluating how "close" her choices are to satisfying the Generalized Axiom of Revealed Preference (GARP), a necessary and sufficient condition for a finite number of choices to be consistent with the maximization of any locally non-satiated utility function. In our framework, the BR ratio can be used to measure the violation of the homogeneous segment assumption. Table 3.3 displays the BR ratios for California and Florida. The BR ratio for π_1 is smaller than the ratio for π_2 in each state; these ratios are roughly the same across states.

Table 3.3. BR ratios for California and Florida.

State	π_1	π_2
California	0.15	0.20
Florida	0.17	0.21

3.7 Concluding Remarks

This chapter is one among pioneering papers attempting to tackle the challenges of performing structural demand analysis with scanner data (see also Burda et al., 2008, 2012, Crawford and Polisson, 2016, Guha and Ng, 2019, Chernozhukov et al., 2020, and Chapter 2). The recent availability of scanner data permits new developments in the analysis of consumer behaviour. Here, I have shown that, by introducing homogeneous segments of consumers, we can consider a model of consumption with non-parametric preferences and infinite-dimensional heterogeneity, not only from a theoretical point-of-view, but also from a practical one. The distribution of individual heterogeneity in the population can be estimated, and the underlying non-parametric preferences can be filtered by using appropriate algorithms.

We developed an analysis for *two goods* for exposition. This feature of our analysis leaves the question: Can the methods developed in this paper be extended to a framework with, say, 100 goods? A completely *unconstrained* non-parametric analysis would encounter the curse of dimensionality. Specifically, we would need to estimate the distribution of the utility function (a non-parametric function with, in this scenario, 100 arguments). This task would be infeasible, even in our big data framework. But, the SARA model with independent taste parameters is a *constrained* non-parametric model. The structure of the SARA model reduces the non-parametric dimension of the problem, making it feasible. Indeed, when taste parameters are independent, we only need to estimate 100 one-dimensional distributions. A similar remark applies to the algorithm used to filter the taste distributions: The two steps based on the bilinear form of the MRS restrictions in a two good setting can be replaced with 100 successive steps based on the multilinear form of MRS restrictions in a 100 good setting, without increasing the numerical complexity.

Many of the results in this paper require taste parameters to be independent, but this requirement can be relaxed. For example, we can always consider a SARA model with the following form:

$$U(x; A) = -\exp(-(A_c + A_1)x_1 - (A_c + A_2)x_2), \tag{3.7.1}$$

where A_c is a common component, and A_j is a good-specific taste parameter, for each $j = 1, 2$. In such a framework, independence between A_c, A_1, and A_2 does not imply independence between the parameters:

$$A_1^* = A_c + A_1 \;\text{ and }\; A_2^* = A_c + A_2, \tag{3.7.2}$$

but it does reduce the dimensionality of the problem: Instead of introducing a joint distribution π on a space of dimension 2, the model only depends on three distributions on a space of dimension 1. This specification avoids the curse of dimensionality. (See Appendix 3.E for a discussion of identification in this case with taste dependence.)

In this paper, consumers are assumed to be rational, and divided into homogeneous segments. Since, in each segment, the demand function can be non-parametrically estimated over a subset of its domain, the analysis can be continued to develop a test of the homogeneity of each segment, or, more generally, a non-parametric method for constructing homogeneous segments.

The approach developed in this paper uses standard ideas from consumer theory to make inference. This approach is appropriate when both quantities and prices have continuous supports. This feature makes this approach valid for some levels of good, consumer, and date aggregation. Hence, this approach can be used for, say, evaluating the effect of alcohol tax on alcohol consumption, but is unreasonable for analyzing how a particular consumer will choose between hundreds of different brands of whiskey. To our knowledge, the tools needed to solve such a problem have not been developed yet.

3.A The Dirichlet Process

In this appendix, I briefly review the definition and properties of the Dirichlet process, and then describe how to simulate from the Dirichlet process (see Ferguson, 1974, Rolin, 1992, Sethuraman, 1994, Lin, 2016, and Li et al., 2019).

3.A.1 Definition and Properties of the Dirichlet Process

For exposition, let us consider the Dirichlet distribution, then the Dirichlet process:

(i) **Dirichlet Distribution:**

Let $D_J(\alpha)$ denote the J-dimensional Dirichlet distribution with density:

$$f_\alpha(q) = \frac{\Gamma\big(\sum_{j=1}^J \alpha_j\big)\prod_{j=1}^J q_j^{\alpha_j}}{\prod_{j=1}^J \Gamma(\alpha_j)}, \tag{3.A.1}$$

for every $q \in [0,1]^J$ such that $\sum_{j=1}^J q_j = 1$, where $\alpha \in \mathbb{R}^J_{++}$ denotes a J-dimensional vector of positive parameters. If a random vector $(Q_1, \dots, Q_J)$ has a Dirichlet distribution $D_J(\alpha)$, then:

$$\mathbb{E}[Q_j] = \bar{\alpha}_j \quad \text{and} \quad V(Q_j) = \frac{\bar{\alpha}_j(1-\bar{\alpha}_j)}{1+\sum_{j=1}^J \alpha_j}, \tag{3.A.2}$$

where $\bar{\alpha}_j = \alpha_j / \sum_{j=1}^J \alpha_j$.

(ii) **Dirichlet Process:**

In the SARA and SSF models, there are two taste parameters, A_1 and A_2. The probability distribution π of (A_1, A_2) is defined on $\mathbb{R}^2_+$. Therefore, in this section, I describe the Dirichlet process in this special case. Let $\mathscr{B}_0$ denote the Borel sets associated with $\mathbb{R}^2_+$, $\mathscr{F}$ denote the set of probability measures defined on $(\mathbb{R}^2_+, \mathscr{B}_0)$, and $\mathscr{B}_1$ denote the σ-algebra consisting of the Borel sets associated with the topology of weak convergence on $\mathscr{F}$. Let μ denote a (deterministic) probability measure defined on $(\mathbb{R}^2_+, \mathscr{B}_0)$, and let c denote a strictly positive scalar. A process G with values in $\mathscr{F}$ is a Dirichlet process with functional parameter μ and scaling parameter c if, for every finite and measurable partition $\{C_1, \ldots, C_J\}$ of $\mathbb{R}^2_+$, the random vector $\left[G(C_1), \ldots, G(C_J)\right]'$ has a J-dimensional Dirichlet distribution given $\alpha = \left[c\mu(C_1), \ldots, c\mu(C_J)\right]'$. There exists a Dirichlet process for every probability measure μ defined on $(\mathbb{R}^2_+, \mathscr{B}_0)$ and scaling parameter c. The distribution of the Dirichlet process is a probability measure defined on $(\mathscr{F}, \mathscr{B}_1)$, whose realizations are almost surely *discrete* probability measures defined on $(\mathbb{R}^2_+, \mathscr{B}_0)$, assigning probability one to the set of all discrete probability measures defined on $(\mathbb{R}^2_+, \mathscr{B}_0)$. The support of the distribution of the Dirichlet process is a set of distributions with support contained in the support of $c\mu$ (Ferguson, 1974). The functional parameter μ (sometimes called the base distribution) can be thought of as the mean of the Dirichlet process—indeed, for any measurable set C in $\mathbb{R}^2_+$, the mean of the Dirichlet distribution in (3.A.2) yields $\mathbb{E}[G(C)] = \mu(C)$. Therefore, in the current framework, μ represents the expected uncertainty on taste parameters. Intuitively, the scaling parameter c describes the "strength" of discretization: When c is large, the realizations of the Dirichlet process are concentrated around μ; loosely, as c tends to infinity, the realizations become "more continuous."

3.A.2 Simulating a Dirichlet Process

A Dirichlet process is easy to simulate given μ and c. There are a number of ways to simulate a realization—this section outlines the *stick-breaking method*, appropriate for drawing under the independence of A_1 and A_2, based on the construction of the Dirichlet process in Sethuraman (1994).

Let $B(\alpha_1, \alpha_2)$ denote the beta distribution with continuous density:

$$f(q) = \frac{\Gamma(\alpha_1 + \alpha_2) q_1^{\alpha_1} q_2^{\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}, \tag{3.A.3}$$

on the simplex $\{(q_1, q_2) \geq 0 : q_1 + q_2 = 1\}$, in which Γ denotes the gamma function, and $\alpha_1, \alpha_2 > 0$ are positive scalar parameters. Under the independence of A_1 and A_2, it is sufficient to be able to make a draw from a Dirichlet process whose realizations are distributions on $[0, \infty)$. Let μ^* and c^* denote the mean and scaling parameter of this Dirichlet process. We can simulate from this process using the following steps:

Step 1: For large $L \geq 1$, independently simulate $W_1, \ldots, W_L \sim B(1, c^*)$.

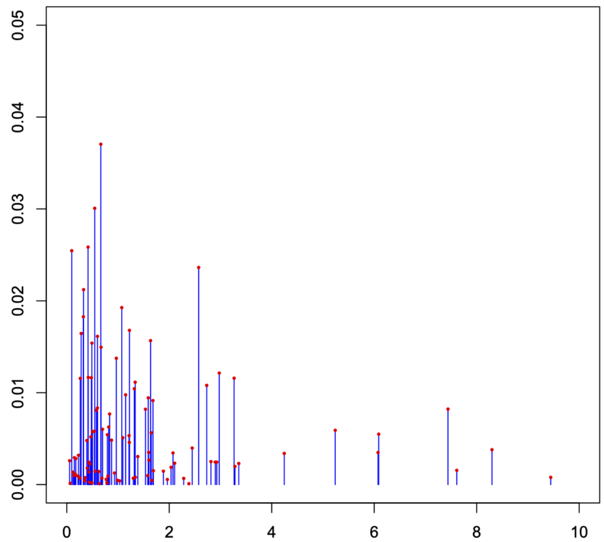

Figure 3.14. A simulated realization from the Dirichlet process given log-normal μ^* with mean 0 and standard deviation 1 (where these parameters are interpreted on the log-scale) and scaling parameter $c = 100$. This realization was simulated using the stick-breaking method given $L = 100$.

Step 2: Compute $W_1^* = W_1$, and:

$$W_\ell^* = W_\ell \prod_{j=1}^{\ell-1}(1 - W_j), \quad \forall \ell = 2, \dots, L. \tag{3.A.4}$$

Step 3: Independently simulate $V_1, \dots, V_L \sim \mu^*$.

Step 4: Define:

$$G(C) = \sum_{\ell=1}^{L} W_\ell^* \delta_{V_\ell}(C), \quad \forall C \subseteq \mathbb{R}_+^2, \tag{3.A.5}$$

where δ_v denotes a point mass at $v \in \mathbb{R}_+^2$.

Theoretically, if we could simulate an infinite number of draws, then this procedure would produce a realization of the Dirichlet process associated with functional parameter μ^* and scaling parameter c^*. Since L is finite, the resulting probability measure G is a *truncated* approximation of a realization of such a process. Figure 3.14 displays a simulated realization from the Dirichlet process given log-normal μ^* with mean 0 and standard deviation 1 (where these parameters are interpreted on the log-scale) and scaling parameter $c = 100$. This realization was simulated using the stick-breaking method given $L = 100$.

3.B Integrability

In each model, the Hessian of the utility function $U(x;\pi)$ is negative definite, implying that the demand function $X(z;\pi)$ is invertible in the second regime (see Chapter 2). Invertibility can also be analyzed using the indifference curves of $U(x;\pi)$, or the Slutsky coefficient $\Delta_x(z)$. As a by-product, this analysis can yield new properties of the moment generating function (MGF).

3.B.1 SARA Model

Suppose A_1 and A_2 are independent. Let Ψ_j denote the Laplace transform of A_j, for $j = 1, 2$. With this notation, we can write:

$$\log U(x;\pi) = \log \Psi_1(x_1) + \log \Psi_2(x_2). \tag{3.B.1}$$

The indifference curve $g_\pi(\cdot, u)$ associated with $U(x;\pi)$ is obtained by solving:

$$\log \Psi_1(x_1) + \log \Psi_2(x_2) = \log u, \tag{3.B.2}$$

for x_2. This procedure leads to:

$$g_\pi(x_1, u) = (\log \Psi_2)^{-1} (\log u - \log \Psi_1(x_1)) . \tag{3.B.3}$$

In general, demand is invertible if the indifference curves are strictly convex such that:

$$\frac{\partial^2 g_\pi(x_1, u)}{\partial x_1^2} > 0, \tag{3.B.4}$$

for every $x_1 > 0$, and every attainable $u < 0$. When preferences are SARA, we obtain:

$$\begin{aligned} &\frac{d^2}{dv^2}\left[(\log \Psi_2)^{-1}\right] (\log u - \log \Psi_1(x_1)) \left(\frac{d \log \Psi_1(x_1)}{dx_1}\right)^2 \\ &-\frac{d}{dv}\left[(\log \Psi_2)^{-1}\right] (\log u - \log \Psi_1(x_1)) \frac{d^2 \log \Psi_1(x_1)}{dx_1^2} > 0, \end{aligned} \tag{3.B.5}$$

for every $x_1 > 0$, and every attainable $u < 0$. These inequalities, involving two MGFs, are always satisfied. Consequently, we have derived a new property of the MGF, as described in the introduction of this appendix.

3.B.2 SSF Model

If preferences are SSF, it is rather challenging to derive a closed-form expression for the indifference curve. We can, instead, write the integrability condition using the condition on the bordered Hessian in Lemma 2.1 in Chapter 2, but, for both brevity and exposition, let us simply restrict our attention to the general specification of utility in the example in Section 3.3.3 and check that integrability holds for any Laplace transform. Because the strict convexity of the indifference curves is equivalent to the strict negativity of the Slutsky coefficient $\Delta_x(z)$, it is sufficient to check whether $\Delta_x(z)$ is strictly negative. We obtain:

$$\Delta_x(z) = \frac{\partial X_1(z;\pi)}{\partial p} + X_1(z;\pi)\frac{\partial X_1(z;\pi)}{\partial y} = -\frac{1}{\lambda p^3}\frac{d}{dv}\left(\frac{d\log\Psi}{dv}\right)^{-1}\left(-\frac{1}{p}\right). \tag{3.B.6}$$

It is sufficient to show that:

$$\frac{d}{dv}\left(\frac{d\log\Psi}{dv}\right)^{-1}\left(-\frac{1}{p}\right), \tag{3.B.7}$$

is strictly positive. To do this, consider the following derivatives:

$$\begin{aligned}\frac{d\log\Psi(v)}{dv} &= -\frac{\mathbb{E}[A_1\exp(-A_1 v)]}{\mathbb{E}[\exp(-A_1 v)]},\\ \text{and}\quad \frac{d^2\log\Psi(v)}{dv^2} &= \frac{\mathbb{E}[A_1^2\exp(-A_1 v)]}{\mathbb{E}[\exp(-A_1 v)]} - \left(\frac{\mathbb{E}[A_1\exp(-A_1 v)]}{\mathbb{E}[\exp(-A_1 v)]}\right)^2 = V_{\tilde{\pi}}(A_1) > 0,\end{aligned} \tag{3.B.8}$$

where the variance is with respect to the transformed density:

$$\frac{\exp(-A_1 v)}{\mathbb{E}[\exp(-A_1 v)]}\pi(v). \tag{3.B.9}$$

Therefore, $\frac{d\log\Psi}{dv}$ is increasing, and so is its inverse $\left(\frac{d\log\Psi}{dv}\right)^{-1}$. Thus, $\Delta_x(z)$ is negative.

3.C Numerical Optimization

The optimization problem for filtering can be written as:

$$\begin{aligned}&\min_{\pi_1,\pi_2}\ (\pi_1-\hat{\pi}_1)'(\pi_1-\hat{\pi}_1)+(\pi_2-\hat{\pi}_2)'(\pi_2-\hat{\pi}_2)\\ &\text{s.t. MRS restrictions (3.6.2)},\ e'\pi_1 = 1,\ \text{and}\ e'\pi_2 = 1,\end{aligned} \tag{3.C.1}$$

where π_1 and π_2 are written on a sufficiently large discrete grid for A_1 and A_2, and $e = (1,\ldots,1)'$. This optimization problem can be difficult due to the dimension of the problem. The objective function is minimized with respect to the total number $2J$ of grid points in π_1 and π_2, which is intentionally chosen to be very large (at least several hundred), and the number of constraints is N_m, where N_m denotes the number of observations with positive consumption x_{it} in segment Λ_m, which is typically around 1,000. Note, $2J$ has to be larger than N_m for identification. Therefore, it is important to find a tractable algorithm for such a problem.

We can use the fact that the MRS restrictions are bilinear in π_1 and π_2. Indeed, these constraints can be written as:

$$A_1(\pi_2)\pi_1 = b_1(\pi_2) \text{ or } A_2(\pi_1)\pi_2 = b_2(\pi_1). \tag{3.C.2}$$

To illustrate, consider the SARA model, and let a_{1j} and a_{2j}, $j = 1, \ldots, J$, denote the locations of the points in the grids for A_1 and A_2, respectively. Moreover, let $\pi_1 = (\pi_{1j})$ and $\pi_2 = (\pi_{2j})$ denote the elementary probabilities on (a_{1j}) and (a_{2j}), respectively. Under the independence of A_1 and A_2, the MRS restrictions have the form:

$$\begin{aligned}&\sum_{j=1}^{J}[\pi_{1j}a_{1j}\exp(-a_{1j}x_{i1t})]\sum_{j=1}^{J}[\pi_{2j}\exp(-a_{2j}x_{i2t})]\\ -p_{it}&\sum_{j=1}^{J}[\pi_{1j}\exp(-a_{1j}x_{i1t})]\sum_{j=1}^{J}[\pi_{2j}a_{2j}\exp(-a_{2j}x_{i2t})] = 0,\end{aligned}$$

for every $i \in \Lambda_m$ and every t with $x_{it} > 0$. The closed-form expressions for $A_1(\pi_2)$, $b_1(\pi_2)$, $A_2(\pi_1)$, and $b_2(\pi_1)$ can be easily deduced. The unit mass restrictions can also be explicitly written as:

$$\sum_{j=1}^{J}\pi_{1j} = 1 \text{ and } \sum_{j=1}^{J}\pi_{2j} = 1.$$

The equivalent expressions in (3.C.2) can be used to solve the optimization problem in (3.C.1) by using a succession of optimization problems with smaller dimensions (see Gouriéroux et al., 1990, and Van Rosen, 2018). Precisely, let $\pi_1(k)$ and $\pi_2(k)$ denote the solutions for π_1 and π_2 at the k^{th} step of the optimization algorithm. Given $\pi_2(k)$, $\pi_1(k+1)$ is defined as the solution to:

$$\min_{\pi_1}\,(\pi_1 - \hat{\pi}_1)'(\pi_1 - \hat{\pi}_1) \text{ s.t. } A_1[\pi_2(k)]\pi_1 = b_1[\pi_2(k)] \text{ and } e'\pi_1 = 1, \tag{3.C.3}$$

and, similarly, $\pi_2(k+1)$ is defined as the solution to:

$$\min_{\pi_2}\,(\pi_2 - \hat{\pi}_2)'(\pi_2 - \hat{\pi}_2) \text{ s.t. } A_2[\pi_1(k+1)]\pi_2 = b_2[\pi_1(k+1)] \text{ and } e'\pi_2 = 1. \tag{3.C.4}$$

If this algorithm numerically converges, then the limit is the solution to the original optimization problem in (3.C.1). Moreover, $\pi_1(k)$ and $\pi_2(k)$ have closed-form solutions:

Proposition 3.8. The solution to (3.C.3) is equal to:

$$\pi_1(k+1) = \hat{\pi}_1 + A_1^*[\pi_2(k)]'\left\{A_1^*[\pi_2(k)]A_1^*[\pi_2(k)]'\right\}^{-1}\left\{b_1^*[\pi_2(k)] - A_1^*[\pi_2(k)]\hat{\pi}_1\right\}, \tag{3.C.5}$$

where A_j^* and b_j^* encompass the MRS constraint and the unit mass contraint together.

Proof. The optimization problem in (3.C.1) is of the following type:

$$\min_{w} (w - w_0)'(w - w_0) \text{ s.t. } Aw = b.$$

Let λ denote a Lagrange multiplier. The first-order conditions are, then:

$$2(w - w_0) - A'\lambda = 0 \text{ and } Aw = b. \tag{3.C.6}$$

The first condition can be written as:

$$w = w_0 + \frac{1}{2}A'\lambda. \tag{3.C.7}$$

By plugging this expression for w into the second condition, we obtain:

$$\frac{\lambda}{2} = (AA')^{-1}(b - Aw_0). \tag{3.C.8}$$

Together, (3.C.7) and (3.C.8) imply:

$$w = w_0 + A'(AA')^{-1}(b - Aw_0).$$

This expression is exactly the form of the solution in the statement of this proposition. □

Remark 3.1. Instead of minimizing the ℓ_2-distance between π_j and $\hat{\pi}_j$, we could use an information criterion, as in Kitamura and Stutzer (1997). However, we would no longer obtain a closed-form solution for $\pi_1(k)$ and $\pi_2(k)$, and we would have to solve a non-linear system in λ with dimension N_m.

Remark 3.2. The inversion of AA' is numerically feasible, but can be made more robust numerically by including a regularization. In particular, it can be replaced with the inversion of $AA' + \varepsilon I$, where $\varepsilon > 0$ is a small regularization parameter. This regularization by shrinkage (see, for example, Ledoit and Wolf, 2004) is preferable to the machine learning practice which replaces AA' with the diagonal matrix made up of the diagonal elements of AA'. It is related to Tikhonov regularization, which can be used to handle the case in which the number of restrictions exceeds $2J$. In practice, it can also be easier to solve the system in (3.C.6), instead of using (3.C.5), and, depending on the case, it can be better to regularize $A'A$, rather than AA'.

Remark 3.3. The optimization problem in (3.C.1) has not explicitly accounted for the positivity of π_1 and π_2. We can incorporate positivity by adjusting after each step of the algorithm.

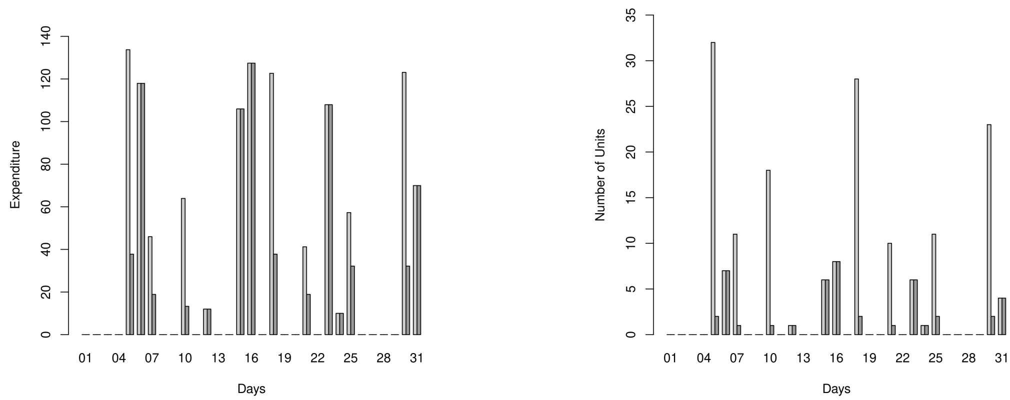

Figure 3.15. **Daily Purchases.** On the left, I illustrate daily expenditure for a single consumer in October of 2016. On the right, we illustrate the number of units purchased by this consumer. Light shaded bars represent all purchases and dark shaded bars represent alcohol-related purchases.

3.D The Nielsen Database

In this appendix, I provide more information about the Nielsen Homescan Consumer Panel (NHCP). First, I describe the individual records, then the representativeness of the restricted sample.

3.D.1 Individual Records

As mentioned in the text, all purchases are continuously recorded by each consumer. The left panel in Figure 3.15 displays the daily (total and alcohol-specific) expenditures of a given consumer in October of 2016. During this month, this consumer purchased 166 units of 97 distinct goods (prior to aggregation). The right panel in Figure 3.15 displays the daily number of units purchased by this consumer. These purchases were all made at three distinct retailers.

3.D.2 Representativeness of Sample

Let us now consider the demographics of the households in the data and compare these demographics with the Current Population Survey (CPS). See Guha and Ng (2019) and Chapter 2 for additional summary statistics.

Table 3.4 gives the distribution of household size in the sample and the CPS. These distributions are similar. The sample has a slightly smaller proportion of households with a single member, and a slightly larger proportion of households with two members. This difference can be explained by single-member households simply buying less alcohol.

Table 3.5 describes the distribution of household income in the sample and the CPS. Once again, these two distributions are quite similar, but the sample has a higher proportion of households earning between $70,000 and $99,999.

Table 3.4. Household size in the sample and in the 2017 Annual Social and Economic Supplement (ASEC) of the CPS. CPS numbers are in thousands.

	Sample		CPS	
Size	Number	Proportion	Number	Proportion
1	5,862	0.2090	35,388	0.2812
2	12,768	0.4554	42,785	0.3400
3	4,121	0.1469	19,423	0.1543
4	3,395	0.1210	16,267	0.1292
5	1,288	0.0459	7,548	0.0599
6	422	0.0150	2,813	0.0223
7+	180	0.0064	1,596	0.0126
Total	28,036	1.0000	125,819	1.0000

Table 3.5. Annual household income in the sample and in the 2017 Annual Social and Economic Supplement (ASEC) of the CPS. CPS numbers are in thousands.

	Sample		CPS	
Income	Number	Proportion	Number	Proportion
Under \$5,000	265	0.0094	4,138	0.0327
\$5,000 to \$9,999	274	0.0097	3,878	0.0307
\$10,000 to \$14,999	638	0.0227	6,122	0.0485
\$15,000 to \$19,999	694	0.0247	5,838	0.0462
\$20,000 to \$24,999	1,147	0.0409	6,245	0.0494
\$25,000 to \$29,999	1,282	0.0457	5,939	0.0470
\$30,000 to \$34,999	1,480	0.0527	5,919	0.0468
\$35,000 to \$39,999	1,432	0.0510	5,727	0.0453
\$40,000 to \$44,999	1,449	0.0516	5,487	0.0434
\$45,000 to \$49,999	1,637	0.0583	5,089	0.0403
\$50,000 to \$59,999	2,878	0.1026	9,417	0.0746
\$60,000 to \$69,999	2,380	0.0848	8,213	0.0650
\$70,000 to \$99,999	6,459	0.2303	19,249	0.1524
\$100,000+	6,021	0.2147	34,963	0.2769
Total	28,036	1.0000	126,224	1.0000

Table 3.6. Age of eldest household head in the sample and the householder in the 2017 Annual Social and Economic Supplement (ASEC) of the CPS. CPS numbers are in thousands.

	Sample		CPS	
Age	Number	Proportion	Number	Proportion
Under 20	4	0.0001	753	0.0059
20 to 24	54	0.0019	5,608	0.0445
25 to 29	476	0.0169	9,453	0.0751
30 to 34	1,201	0.0428	10,594	0.0842
35 to 39	1,817	0.0648	10,651	0.0846
40 to 44	1,893	0.0675	10,571	0.0840
45 to 49	2,398	0.0855	11,115	0.0883
50 to 54	3,058	0.1090	12,180	0.0968
55 to 64	7,869	0.2806	23,896	0.1899
65 to 74	6,507	0.2320	17,551	0.1394
75+	2,759	0.0984	13,448	0.1068
Total	28,036	1.0000	125,819	1.0000

Table 3.6 gives the distribution of the age of the eldest head of the household in our sample and the age of the householder in the CPS. There is no direct comparison for these statistics, as the eldest head may differ from the householder. This aspect of the data can explain why our sample seems to be older than the general population.

There may also exist another source of non-representativeness: A consumer might behave differently because she is being observed. For example, she might increase her expenditure to give the impression that she is richer. This type of behaviour can be observed when the period of observation is short, but is not usually sustainable in the long term. This effect should be negligible over the four months considered in the illustration in Section 3.6.

3.E SARA Model with Taste Dependence

Consider the SARA model with taste dependence described in Section 1.6. In this model, dependence is introduced using a "common" stochastic taste parameter such that:

$$U(x;\pi) = -\mathbb{E}_\pi\big[\exp(-(A_c + A_1)x_1 - (A_c + A_2)x_2)\big], \tag{3.E.1}$$

where A_c, A_1, A_2 are independent non-negative taste parameters with distributions π_c, π_1, and π_2. This utility function can be written in terms of the Laplace transforms of these taste parameters:

$$U(x;\pi) = -\Psi_c(x_1 + x_2)\Psi_1(x_1)\Psi_2(x_2), \tag{3.E.2}$$

where Ψ_c, Ψ_1, and Ψ_2 are the Laplace transforms of A_c, A_1, and A_2. Therefore, its marginal rate of substitution has the form:

$$\mathrm{MRS}(x;\pi) = \frac{d \log \Psi_c(x_1+x_2)/dx + d \log \Psi_1(x_1)/dx}{d \log \Psi_c(x_1+x_2)/dx + d \log \Psi_2(x_2)/dx}. \tag{3.E.3}$$

Below, I consider the possibility to identify the distributions of A_c, A_1, and A_2 from the knowledge of the utility function (not observable), and then from the knowledge of the marginal rate of substitution (which can be obtained by inverting the demand).

3.E.1 Identification from the Utility Function

I first ask whether the knowledge of the utility function is equivalent to the knowledge of the distributions π_c, π_1, and π_2. If the utility function $U(x;\pi)$ is known, then:

$$\log(-U(x;\pi)) = \log \Psi_c(x_1+x_2) + \log \Psi_1(x_1) + \log \Psi_2(x_2), \tag{3.E.4}$$

is known. By taking the cross-derivative of this expression with respect to x_1 and x_2, we can also obtain knowledge of:

$$\frac{d^2 \log \Psi_c(x_1+x_2)}{dx^2}, \tag{3.E.5}$$

for any $x_1, x_2 > 0$. Consequently, $\log \Psi_c(x)$ is known up to an affine function $\alpha x + c$. Moreover, since $\log \Psi_c(0) = 0$, we can identify $\log \Psi_c(x)$ up to a multiplicative factor α. Equivalently, A_c can be replaced with $A_c - \alpha$, and A_j^* can be replaced with $A_j^* + \alpha$, for $j = 1, 2$, without changing the utility function. This reasoning leads to the following result:

Proposition 3.9. If preferences are SARA with a common stochastic taste parameter, A_c, A_1, and A_2 are independent, and the distributions π_c, π_1, and π_2 have support $(0, \infty)$, then these distributions are identified from the observation of utility $U(x;\pi)$.

Proof. If π_c, π_1, and π_2 have support $(0, \infty)$, then $\alpha = 0$. The identification follows. □

Proposition 3.9 shows that a condition on the supports of the taste distributions is needed for identification.

3.E.2 Identification Issue

Let us now consider the possibility of two distinct sets of taste distributions resulting in the same preference ordering (equivalently, the same marginal rate of substitution).

Proposition 3.10. If preferences are SARA with a common stochastic taste parameter, and A_c, A_1, and A_2 are independent, then (Ψ_c, Ψ_1, Ψ_2) and $(\Psi_c^\nu, \Psi_1^\nu, \Psi_2^\nu)$ lead to the same preference ordering, for all positive scalars $\nu > 0$.

Proof. The proof is a direct consequence of the expression for $\mathrm{MRS}(x;\pi)$ in (3.E.3). $\square$

This type of identification issue has already been encountered in the SARA model with independent taste parameters, as described in Section 3.5.1. It is not surprising that we have a similar result in this model.

3.E.3 Recursive Case

We are left with the question: Is the identification issue described above the only type of issue that we encounter in this model? First, let us consider the case in which $A_1 = 0$. In this case, the total taste parameter for good 1 is A_c, and the total taste parameter for good 2 is $A_c + A_2$. Therefore, the consumer's risk aversion for drinks with high ABV is systematically larger than her risk aversion for drinks with low ABV.

Proposition 3.11. If preferences are SARA with a common stochastic taste parameter, $A_1 = 0$, A_c and A_2 are independent, and the distributions π_c and π_2 have support $(0, \infty)$, then these distributions are identified up to a power transform of their Laplace transforms.

Proof. If $A_1 = 0$, then the knowledge of the MRS implies the knowledge of:

$$\frac{d \log \Psi_c(x_1 + x_2)/dx}{d \log \Psi_2(x_2)/dx}. \tag{3.E.6}$$

By considering $x_2 = 0$, we have to solve the equation:

$$\frac{d \log \Psi_c(x_1)}{dx} = \frac{d \log \Psi_2(0)/dx}{d \log \Psi_2^*(0)/dx} \cdot \frac{d \log \Psi_c^*(x_1)}{dx} \tag{3.E.7}$$

Therefore, there is a positive scalar ν such that:

$$\frac{d \log \Psi_c(x_1)}{dx} = \nu \frac{d \log \Psi_c^*(x_1)}{dx}, \tag{3.E.8}$$

and the result follows by integration, using $\Psi_c(0) = \Psi_c^*(0) = 1$.

$\square$

3.E.4 General Case

Let us now consider the general case. The proof of the following result uses the limiting behaviour of the Laplace transform of a positive random variable as x tends to infinity.

Proposition 3.12. If preferences are SARA with a common stochastic taste parameter, A_c, A_1, and A_2 are independent, and the distributions π_c, π_1, and π_2 have support $(0, \infty)$, then (Ψ_c, Ψ_1, Ψ_2) and

$(\Psi_c^*, \Psi_1^*, \Psi_2^*)$ lead to the same preference ordering if, and only if, for some $\nu > 0$, we have:

$$\Psi_c = (\Psi_c^*)^\nu, \;\; \Psi_1 = (\Psi_1^*)^\nu, \;\; \text{and} \;\; \Psi_2 = (\Psi_2^*)^\nu. \tag{3.E.9}$$

Proof. Consider the following two steps to identification:

(i) For any positive variable A, the logarithmic derivative of its Laplace transform $\Psi(x)$ equals:

$$\frac{d \log \Psi(x)}{dx} = -\frac{\mathbb{E}[A \exp(-Ax)]}{\mathbb{E}[\exp(-Ax)]} = \mathbb{E}_{Q_x}[A], \tag{3.E.10}$$

where Q_x is the deformed probability distribution with density:

$$\frac{\exp(-Ax)}{\mathbb{E}[\exp(-Ax)]}, \tag{3.E.11}$$

with respect to the distribution of A. This deformed distribution tends to the positive point mass at 0, and $d \log \Psi(x)/dx$ tends to 0, when x tends to infinity, under the assumption that A has full support.[17] This result can be applied to A_c, A_1, and A_2.

(ii) We can use the MRS in (3.E.3) to identify:

$$\frac{d \log \Psi_1(x_1)/dx - d \log \Psi_2(x_2)/dx}{d \log \Psi_c(x_1 + x_2)/dx}. \tag{3.E.12}$$

Therefore, we can look for the solution(s) to the equality:

$$\begin{aligned} &\left(\frac{d \log \Psi_1(x_1)}{dx} - \frac{d \log \Psi_2(x_2)}{dx}\right) \frac{d \log \Psi_c^*(x_1 + x_2)}{dx} \\ =&\left(\frac{d \log \Psi_1^*(x_1)}{dx} - \frac{d \log \Psi_2^*(x_2)}{dx}\right) \frac{d \log \Psi_c(x_1 + x_2)}{dx}. \end{aligned} \tag{3.E.13}$$

Let us now assume that x_2 tends to infinity, and denote $z = x_1 + x_2$. We obtain:

$$\frac{d \log \Psi_1(x_1)}{dx} \frac{d \log \Psi_c^*(z)}{dx} \simeq \frac{d \log \Psi_1^*(x_1)}{dx} \frac{d \log \Psi_c(z)}{dx}, \tag{3.E.14}$$

for any x_1 and large z. Consequently:

$$\frac{d \log \Psi_1(x_1)/dx}{d \log \Psi_1^*(x_1)/dx} \sim \frac{d \log \Psi_c(z)/dx}{d \log \Psi_c^*(z)/dx}. \tag{3.E.15}$$

Since the left-hand side of this equivalence is fixed for z tending to infinity, this limit exists:

$$\lim_{z \to \infty} \frac{d \log \Psi_c(z)/dx}{d \log \Psi_c^*(z)/dx} = \frac{d \log \Psi_1(x_1)/dx}{d \log \Psi_1^*(x_1)/dx}. \tag{3.E.16}$$

[17] Without this assumption, we get $\lim_{x\to\infty} \frac{d \log \Psi(x)}{dx} = \operatorname{ess\,inf} A$, where ess inf denotes the essential infimum of the distribution of A (see Theorem 13.2.5 and Remark 13.3 in Polyanskiy and Wu, 2017).

Moreover, by definition, this limit is independent of x_1. We deduce that:

$$\frac{d \log \Psi_1(x_1)}{dx} = \nu \frac{d \log \Psi_1^*(x_1)}{dx}, \tag{3.E.17}$$

for some positive scalar ν.

(iii) The same reasoning can be used as x_1 tends to infinity. This procedure yields:

$$\nu = \lim_{z \to \infty} \frac{d \log \Psi_c(z)/dx}{d \log \Psi_c^*(z)/dx} = \frac{d \log \Psi_2(x_2)/dx}{d \log \Psi_2^*(x_2)/dx}. \tag{3.E.18}$$

Therefore, we obtain:

$$\frac{d \log \Psi_c(x)}{dx} = \nu \frac{d \log \Psi_c^*(x)}{dx} \quad \text{and} \quad \frac{d \log \Psi_j(x)}{dx} = \nu \frac{d \log \Psi_j^*(x)}{dx}, \tag{3.E.19}$$

for $j = 1, 2$, and the result in this proposition follows by integration.

□

Bibliography

Afriat, S. (1967): "The Construction of Utility Functions from Expenditure Data," *International Economic Review*, 8, 67–77.

Allen, R. (1936): "Professor Slutsky's Theory of Consumers' Choice," *The Review of Economic Studies*, 3, 120–129.

Allen, R. and J. Rehbeck (2019): "Assessing Misspecification and Aggregation of Structured Preferences," *Working Paper*.

Alley, A., D. Ferguson, and K. Stewart (1992): "An Almost Ideal Demand System for Alcoholic Beverages in British Columbia," *Empirical Economics*, 17, 401–418.

Andrews, D. (1995): "Non-Parametric Kernel Estimation for Semi-Parametric Models," *Econometric Theory*, 11, 560–596.

Andrews, D. and X. Shi (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81, 609–666.

Arrow, K. and A. Enthoven (1961): "Quasi-Concave Programming," *Econometrica*, 29, 770–800.

Asmussen, S., J. Jensen, and L. Rojas-Nandayapa (2016): "On the Laplace Transform of the Lognormal Distribution," *Methodology and Computing in Applied Probability*, 18, 441–458.

Aussenberg, R. (2018): "Errors and Fraud in the Supplemental Nutrition Assistance Program (SNAP)," *Congressional Research Service*.

Banks, J., R. Blundell, and A. Lewbel (1997): "Quadratic Engel Curves and Consumer Demand," *The Review of Economics and Statistics*, 79, 527–539.

Bansal, R. and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *The Journal of Finance*, 59, 1481–1509.

Barrodale, I. and F. Roberts (1974): "Solution of an Overdetermined System of Equations in the ℓ_1 Norm," *Communications of the ACM*, 17, 319–320.

Barten, A. (1968): "Estimating Demand Equations," *Econometrica*, 36, 213–251.

Barten, A. and V. Böhm (1993): "Consumer Theory," in *Handbook of Mathematical Economics*, vol. 2, chap. 9, 381–429.

Beckert, W. and R. Blundell (2008): "Heterogeneity and the Non-Parametric Analysis of Consumer Choice: Conditions for Invertibility," *The Review of Economic Studies*, 75, 1069–1080.

Bengs, V. and H. Holzmann (2019): "Adaptive Confidence Sets for Kink Estimation," *Electronic Journal of Statistics*, 13, 1523–1579.

Bergstra, J. and Y. Yoshua (2012): "Random Search for Hyper-Parameter Optimization," *Journal of Machine Learning Research*, 13, 281–305.

Billingsley, P. (1999): *Convergence of Probability Measures*, New York: Wiley, 2nd ed.

Blaschke, W. and G. Pick (1916): "Distanzschätzungen im Funktionenraum II," *Mathematische Annalen*, 77, 277–302.

Blomquist, S., A. Kumar, C. Liang, and W. Newey (2015): "Individual Heterogeneity, Nonlinear Budget Sets, and Taxable Income," *CESifo Working Paper*.

Blundell, R., M. Browning, and I. Crawford (2003): "Nonparametric Engel Curves and Revealed Preference," *Econometrica*, 71, 205–240.

Blundell, R., J. Horowitz, and M. Parey (2017): "Nonparametric Estimation of a Nonseparable Demand Function under the Slutsky Inequality Restriction," *The Review of Economics and Statistics*, 99, 291–304.

Blundell, R., L. Pistaferri, and I. Preston (2006): "Imputing Consumption in the PSID using Food Demand Estimates from the CEX," *Working Paper*.

Bolduc, D., R. Laferrière, and G. Santarossa (1992): "Spatial Autoregressive Error Components in Travel Flow Models," *Regional Science and Urban Economics*, 22, 371–385.

Boucher, V. and I. Mourifié (2017): "My Friend Far, Far Away: A Random Field Approach to Exponential Random Graph Models," *Econometrics Journal*, 20, S14–S46.

Bowles, S. (1998): "Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions," *Journal of Economic Literature*, 36, 75–111.

Brown, B. and M. Walker (1989): "The Random Utility Hypothesis and Inference in Demand Systems," *Econometrica*, 57, 815–829.

Brown, D. and R. Matzkin (1995): "The Random Utility Model from Data on Consumer Demand," *Discussion Paper, Yale University.*

Burda, M., M. Harding, and J. Hausman (2008): "A Bayesian Mixed Logit-Probit Model for Multinomial Choice," *Journal of Econometrics*, 147, 232–246.

——— (2012): "A Poisson Mixture Model of Discrete Choice," *Journal of Econometrics*, 166, 184–203.

Cataldo, A. and A. Cohen (2018): "Reversing the Similarity Effect: The Effect of Presentation Format," *Cognition*, 175, 141–156.

Chen, L. (1997): "Multivariate Regression Splines," *Computational Statistics and Data Analysis*, 26, 71–82.

Chen, Z. (1993): "Fitting Multivariate Regression Functions by Interaction Spline Models," *Journal of the Royal Statistical Society*, 55, 473–491.

Cheng, M. and M. Raimondo (2008): "Kernel Methods for Optimal Change-Points Estimation in Derivatives," *Journal of Computational and Graphical Statistics*, 17, 56–75.

Chernozhukov, V., J. Hausman, and W. Newey (2020): "Demand Analysis with Many Prices," *Working Paper.*

Child Nutrition Act (1966): *Amended Through P.L. 92-433.*

Christensen, L., D. Jorgenson, and L. Lau (1975): "Transcendental Logarithmic Utility Functions," *The American Economic Review*, 65, 367–383.

Christensen, O. and S. Goh (2017): "Construction of Scaling Partitions of Unity," *Frontiers in Applied Mathematics and Statistics*, 3.

Clement, E., C. Gouriéroux, and A. Monfort (2000): "Econometric Specification of the Risk-Neutral Valuation Model," *Journal of Econometrics*, 94, 117–143.

Cleveland, W. (1979): "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74, 829–836.

Code of Federal Regulations (2018): *Title 7.*

Connecticut State Department of Public Health (2020): "eWIC: Frequently Asked Questions," Accessed at https://portal.ct.gov/DPH/WIC/eWIC-Frequently-Asked-Questions on May 2, 2020.

Crawford, I. (2019): "Nonparametric Analysis of Labour Supply using Random Fields," *Working Paper.*

Crawford, I. and M. Polisson (2016): "Demand Analysis with Partially Observed Prices," *Institute of Fiscal Studies*.

Cressie, N. (1993): *Statistics for Spatial Data*, New York: Wiley, 2nd ed.

Crosetto, P. and A. Gaudeul (2016): "A Monetary Measure of the Strength and Robustness of the Attraction Effect," *Economics Letters*, 149, 38–43.

de Clippel, G. and K. Rozen (2020): "Relaxed Optimization: ε-Rationalizability and the FOC-Departure Index in Consumer Theory," *Working Paper 2020-07, Brown University*.

Dean, M. and D. Martin (2016): "Measuring Rationality with the Minimum Cost of Revealed Preference Violations," *The Review of Economics and Statistics*, 98, 524–534.

Deaton, A. and J. Muellbauer (1980a): "An Almost Ideal Demand System," *The American Economic Review*, 70, 312–326.

——— (1980b): *Economics and Consumer Behavior*, New York: Cambridge University Press.

Deb, R., Y. Kitamura, J. Quah, and J. Stoye (2018): "Revealed Price Preference: Theory and Empirical Analysis," *Working Paper*.

Department of Defense Appropriations Act (2004-2017): *P.L. 111-118*.

Dette, H., S. Hoderlein, and N. Neumeyer (2016): "Testing Multivariate Economic Restrictions Using Quantiles: The Example of Slutsky Negative Semidefiniteness," *Journal of Econometrics*, 191, 129–144.

Dieudonné, J. (1960): *Foundations of Modern Analysis*, New York: Academic Press.

Dobronyi, C. and C. Gouriéroux (2020): "Stochastic Revealed Preference: A Non-Parametric Analysis," *Working Paper*.

Domencich, T. and D. McFadden (1975): *Urban Travel Demand: A Behavioral Analysis*, New York: North-Holland Publishing Co.

Echenique, F., S. Lee, and M. Shum (2011): "The Money Pump as a Measure of Revealed Preference Violations," *Journal of Political Economy*, 119, 1201–1223.

Fan, J. and I. Gijbels (1992): "Variable Bandwidth and Local Linear Regression Smoothers," *The Annals of Statistics*, 20, 2008–2036.

Feller, W. (1968): *An Introduction to Probability Theory and its Applications*, vol. 2, Wiley.

Ferguson, T. (1974): "Prior Distributions on Spaces of Probability Measures," *The Annals of Statistics*, 2, 615–629.

Food and Nutrition Act (2008): *Amended Through P.L. 113–128.*

Food and Nutrition Service (2006-2017): "Characteristics of Supplemental Nutrition Assistance Program Households," *Annual Report.*

——— (2015): "WIC Food Packages: Maximum Monthly Allowances," Accessed at https://www.fns.usda.gov/wic/wic-food-packages-maximum-monthly-allowances on May 2, 2020.

——— (2017a): "Fiscal Year 2018 Cost-of-Living Adjustments," Accessed at http://www.fns-prod.azureedge.net/sites/default/files/snap/SNAP_Fiscal_Year_2018_Cost_of_Living_Adjustments.pdf on December 14, 2018.

——— (2017b): "A Short History of SNAP," Accessed at http://www.fns.usda.gov/snap/short-history-snap on June 4, 2018.

——— (2018a): "Am I Eligible for SNAP?" Accessed at http://www.fns.usda.gov/snap/eligibility on June 4, 2018.

——— (2018b): "Supplemental Nutrition Assistance Program: Monthly Issuance Schedule for All States and Territories," Accessed at https://fns-prod.azureedge.net/sites/default/files/snap/Monthly-Issuance-Schedule-All-States.pdf on December 14, 2018.

——— (2018c): "Supplemental Nutrition Assistance Program Participation and Costs," Accessed at https://fns-prod.azureedge.net/sites/default/files/pd/SNAPsummary.pdf on December 19, 2018.

——— (2018d): "What Can SNAP Buy?" Accessed at https://fns.usda.gov/snap/eligible-food-items on June 4, 2018.

——— (2020a): "2018-2020 WIC Shopping Guide," Accessed at https://www.health.state.mn.us/docs/people/wic/vendor/fpchng/shopguide.pdf on May 2, 2020.

——— (2020b): "The Special Supplemental Nutrition Program for Women, Infants and Children (WIC Program)," Accessed at https://fns-prod.azureedge.net/sites/default/files/wic/wic-fact-sheet.pdf on May 2, 2020.

——— (2020c): "WIC EBT Activities," Accessed at https://www.fns.usda.gov/wic/wic-ebt-activities on May 2, 2020.

——— (2020d): "WIC Program Participation Costs," Accessed at https://fns-prod.azureedge.net/sites/default/files/resource-files/wisummary-4.pdf on May 2, 2020.

Food Stamp Act (1977a): *Amended Through P.L. 108-269.*

——— (1977b): *P.L. 88-525.*

Gale, D. and H. Nikaidô (1965): "The Jacobian Matrix and Global Univalence of Mappings," *Mathematische Annalen*, 159, 81–93.

Gasser, T. and H. Müller (1984): "Estimating Regression Functions and Their Derivatives by the Kernel Method," *Scandinavian Journal of Statistics*, 11, 171–185.

Gelfand, I. and N. Vilenkin (1964): *Generalized Functions*, Academic Press.

Geweke, J. (2012): "Nonparametric Bayesian Modelling of Monotone Preferences for Discrete Choice Experiments," *The Journal of Econometrics*, 171, 185–204.

Gijbels, I. and A. Goderniaux (2004a): "Bootstrap Test for Change-Points in Nonparametric Regression," *Journal of Nonparametric Statistics*, 16, 591–611.

——— (2004b): "Data-Driven Discontinuity Detection in Derivatives of a Regression Function," *Communications in Statistics*, 33, 851–871.

Gikhman, I. and A. Skorokhod (1966): "On the Densities of Probability Measures in Function Spaces," *Russian Mathematical Surveys*, 21, 83–156.

Ginsberg, W. (1973): "Concavity and Quasiconcavity in Economics," *Journal of Economic Theory*, 6, 596–605.

Goldstein, R. (2000): "The Term Structure of Interest Rates as a Random Field," *The Review of Financial Studies*, 13, 365–384.

Gorman, W. (1953): "Community Preference Fields," *Econometrica*, 21, 63–80.

——— (1961): "On a Class of Preference Fields," *Metroeconomica*, 13, 53–56.

Gouriéroux, C. and A. Monfort (1995): *Statistics and Econometric Models*, vol. 2, New York: Cambridge University Press.

——— (1996): *Simulation-Based Econometric Methods*, New York: Oxford University Press.

Gouriéroux, C., A. Monfort, and E. Renault (1990): "Bilinear Constraints: Estimation and Tests," *Essays in Honor of Edmond Malinvaud, Empirical Economics, MIT Press*, 166–191.

Government Accountability Office (2006): "Food Stamp Trafficking: FNS Could Enhance Program Integrity by Better Targeting Stores Likely to Traffic and Increasing Penalties," *Report to Congressional Committees.*

Grandmont, J.-M. (1992): "Transformations of the Commodity Space, Behavioral Heterogeneity, and the Aggregation Problem," *The Journal of Economic Theory*, 57, 1–35.

GRANT, S. (1995): “A Strong (Ross) Characterization of Multivariate Risk Aversion,” *Theory and Decision*, 38, 131–152.

GUHA, R. AND S. NG (2019): “A Machine Learning Analysis of Seasonal and Cyclical Sales in Weekly Scanner Data,” *NBER 25899*.

HAAG, B., S. HODERLEIN, AND K. PENDAKUR (2009): “Testing and Imposing Slutsky Symmetry in Nonparametric Demand Systems,” *Journal of Econometrics*, 153, 33–50.

HALEVY, Y. AND V. FELTKAMP (2005): “A Bayesian Approach to Uncertainty Aversion,” *Review of Economic Studies*, 72, 449–466.

HALLIN, M., Z. LU, AND K. YU (2009): “Local Linear Spatial Quantile Regression,” *Bernoulli*, 15, 659–686.

HAN, S., R. MESQUITA, T. BUSCH, AND M. PUTT (2014): “A Method for Choosing the Smoothing Parameter in a Semi-Parametric Model for Detecting Change-Points in Blood Flow,” *Journal of Applied Statistics*, 41, 26–45.

HARRELL, F. (2001): *Regression Modeling Strategies*, New York: Springer, 2nd ed.

HASTINGS, J., R. KESSLER, AND J. SHAPIRO (2020): “The Effect of SNAP on the Composition of Purchased Foods: Evidence and Implications,” *American Economic Journal, Forthcoming*.

HAUSMAN, J. AND W. NEWEY (2016a): “Individual Heterogeneity and Average Welfare,” *Econometrica*, 84, 1225–1248.

——— (2016b): “Individual Heterogeneity and Average Welfare,” *Econometrica*, 84, 1225–1248.

HEALTH AND HUMAN SERVICES DEPARTMENT (2004-2017): “Annual Update of the HHS Poverty Guidelines,” *Washington, DC.*

HEALTHY HUNGER-FREE KIDS ACT (2010): *P.L. 111-296*.

HEYDE, C. (1963): “On a Property of the Lognormal Distribution,” *Journal of the Royal Statistical Society Series B: Methodological*, 29, 16–18.

HILDENBRAND, W. (1994): *Market Demand: Theory and Empirical Evidence*, Princeton: Princeton University Press.

HOSOYA, Y. (2013): “Measuring Utility from Demand,” *Journal of Mathematical Economics*, 49, 82–96.

——— (2016): “On First-Order Partial Differential Equations: An Existence Theorem and its Applications,” *Advances in Mathematical Economics*, 20, 77–87.

——— (2020): "The Recoverability Revisited," *Journal of Mathematical Economics*, 90, 31–41.

Houthakker, H. (1950): "Revealed Preference and the Utility Function," *Economica*, 17, 159–174.

Houtman, M. and J. Maks (1985): "Determining all Maximal Data Subsets Consistent with Revealed Preference," *Kwantitatieve Methoden*, 19, 89–104.

Huber, J., J. Payne, and C. Puto (1982): "Let's be Honest about the Attraction Effect," *Journal of Marketing Research*, 51, 520–525.

Huh, J. and K. Carriere (2002): "Estimation of Regression Functions with a Discontinuity in a Derivative with Local Polynomial Fits," *Statistics and Probability Letters*, 56, 329–343.

Hurwicz, L. and H. Uzawa (1971): "On the Integrability of Demand Functions," in *Preferences, Utility, and Demand: A Minnesota Symposium*, New York, chap. 6, 114–148.

Imbens, G. and W. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77, 1481–1512.

Izhakian, Y. (2020): "A Theoretical Foundation of Ambiguity Measurement," *Journal of Economic Theory*, 187, 1–43.

Johansen, L. (1969): "On the Relationships Between Some Systems of Demand Functions," *Liiketaloudellinen Aikakauskirga*, 1, 30–41.

Johansen, S. (1974): "The Extremal Convex Functions," *Mathematica Scandinavica*, 34, 61–68.

Kadane, J. (1971): "Comparison of k-Class Estimators when the Disturbances are Small," *Econometrica*, 39, 723–737.

Karni, E. (1979): "On Multivariate Risk Aversion," *Econometrica*, 47, 1391–1401.

——— (1983): "On the Correspondence Between Multivariate Risk Aversion and Risk Aversion with State-Dependent Preferences," *Journal of Economic Theory*, 30, 230–242.

Katzner, D. (1968): "A Note on the Differentiability of Consumer Demand Functions," *Econometrica*, 36, 415–418.

Kennedy, D. (1994): "The Term Structure of Interest Rates as a Gaussian Random Field," *Mathematical Finance*, 4, 247–258.

Kitamura, Y. and J. Stoye (2018): "Nonparametric Analysis of Random Utility Models," *Econometrica*, 86, 1883–1909.

Kitamura, Y. and M. Stutzer (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65, 861–874.

Klotsche, J. and A. Gloster (2012): "Estimating a Meaningful Point of Change: A Comparison of Exploratory Techniques Based on Nonparametric Regression," *Journal of Educational and Behavioural Statistics*, 37, 579–600.

Koenker, R. (2005): *Quantile Regression*, Cambridge University Press.

Koenker, R. and G. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

Koenker, R. and V. D'Orey (1987): "Algorithm AS 299: Computing Regression Quantiles," *Journal of the Royal Statistical Society*, 36, 383–393.

——— (1994): "Remark AS R92: A Remark on Algorithm AS 229: Computing Dual Regression Quantiles and Regression Rank Scores," *Journal of the Royal Statistical Society*, 43, 410–414.

Koenker, R., P. Ng, and S. Portnoy (1994): "Quantile Smoothing Splines," *Biometrika*, 81, 673–680.

Koopmans, T. and O. Reiersøl (1950): "The Identification of Structural Characteristics," *The Annals of Mathematical Statistics*, 21, 165–181.

Kotz, S., N. Balakrishnan, and N. Johnson (2000): *Continuous Multivariate Distributions, Volume 1: Models and Applications*, New York: Wiley, 2nd ed.

Landais, C. (2015): "Assessing the Welfare Effects of Unemployment Benefits Using the Regression Kink Design," *American Economic Journal: Economic Policy*, 7, 243–278.

Ledoit, O. and M. Wolf (2004): "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices," *Journal of Multivariate Analysis*, 88, 365–411.

Leroux, A. (1987): "Preferences and Normal Goods: A Sufficient Condition," *Journal of Economic Theory*, 43, 192–199.

Leung, M. (2015): "Two-Step Estimation of Network-Formation Models with Incomplete Information," *Journal of Econometrics*, 188, 182–195.

Lewbel, A. (1995): "Consistent Nonparametric Hypothesis Tests with an Application to Slutsky Symmetry," *Journal of Econometrics*, 67, 379–401.

——— (2001): "Demand Systems with and without Errors," *The American Economic Review*, 91, 611–618.

Li, Y., E. Schofield, and M. Gönen (2019): "A Tutorial on Dirichlet Process Mixture Modeling," *Journal of Mathematical Psychology*, 91, 128–144.

Lin, J. (2016): "On the Dirichlet Distribution," *Queens University, Kingston.*

Lin, X. (2005): "Peer Effects and Student Academic Achievement: An Application of Spatial Autoregressive Model with Group Unobservables," *Unpublished Manuscript, Department of Economics, Ohio State University*.

——— (2010): "Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables," *Journal of Labor Economics*, 28.

Lindvall, T. (1973): "Weak Convergence of Probability Measures and Random Functions in the Function Space $D[0, \infty)$," *The Journal of Applied Probability*, 10, 109–121.

Lu, Z. (1996): "Multivariate Locally Weighted Polynomial Fitting and Partial Derivative Estimation," *The Journal of Multivariate Analysis*, 59, 187–205.

Manning, W., L. Blumberg, and L. Moulton (1995): "The Demand for Alcohol: The Differential Response to Price," *Journal of Health Economics*, 14, 123–148.

Mas-Colell, A. (1977): "The Recoverability of Consumers' Preferences from Market Demand Behavior," *Econometrica*, 45, 1409–1430.

——— (1979): "Homeomorphisms of Compact, Convex Sets and the Jacobian Matrix," *SIAM Journal of Mathematical Analysis*, 10, 1105–1109.

——— (1985): *The Theory of General Economic Equilibrium: A Differential Approach*, New York: Cambridge University Press.

Mas-Colell, A., M. Whinston, and J. Green (1995): *Microeconomic Theory*, New York: Oxford University Press, 1st ed.

Masry, E. (1996): "Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates," *The Journal of Time Series Analysis*, 17, 571–599.

Matzkin, R. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339–1375.

McFadden, D. (1981): "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski and D. McFadden, Cambridge: MIT Press, chap. 5, 198–272.

——— (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica*, 57, 995–1026.

——— (2005): "Revealed Stochastic Preference," *Economic Theory*, 26, 245–264.

Miller, Z., M. Levin, and R. Valentine (2017): "SNAP Trafficking Calculator: A Database Solution for Improving Program Integrity," *IMPAQ International*.

MOSCHINI, G. (1998): "The Semiflexible Almost Ideal Demand System," *European Economic Review*, 42, 349–364.

MUELLBAUER, J. (1976): "Community Preferences and the Representative Consumer," *Econometrica*, 44, 525–543.

MUELLER, H. (1992): "Change-Points in Nonparametric Regression Analysis," *Annals of Statistics*, 20, 737–761.

MUELLER, H. AND K. SONG (1997): "Two-Stage Change-Point Estimators in Smooth Regression Models," *Statistics and Probability Letters*, 34, 323–335.

NADARAYA, E. (1964): "On Estimating Regression," *Theory of Probability and its Applications*, 9, 141–142.

NAVARRO, D., T. GRIFFITHS, M. STEYVERS, AND M. LEE (2006): "Modeling Individual Differences Using Dirichlet Processes," *Journal of Mathematical Psychology*, 50, 101–122.

NG, S. (2017): "Opportunities and Challenges: Lessons from Analyzing Terabytes of Scanner Data," in *Advances in Economics and Econometrics*, ed. by B. Honoré, A. Pakes, and L. Samuelson, Cambridge University Press, vol. 2.

NOGUCHI, T. AND N. STEWART (2014): "In the Attraction, Compromise, and Similarity Effects, Alternatives are Repeatedly Compared in Pairs on Single Dimensions," *Cognition*, 132, 44–56.

ORNSTEIN, L. AND G. UHLENBECK (1930): "On the Theory of Brownian Motion," *Physical Review*, 36, 823–841.

PAKES, A. AND D. POLLARD (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057.

POLYANSKIY, Y. AND Y. WU (2017): "Lecture Notes on Information Theory," *Yale University*.

RASMUSSEN, C. AND C. WILLIAMS (2006): *Gaussian Processes for Machine Learning*, Cambridge: MIT Press, 1st ed.

RICHARD, S. (1975): "Multivariate Risk Aversion, Utility Independence and Separable Utility Functions," *Management Science*, 22, 12–21.

RILSTONE, P. (1985): "Nonparametric Partial Derivative Estimation," *University of Western Ontario*.

ROBINSON, P. (1983): "Nonparametric Estimators for Time Series," *Journal of Time Series Analysis*, 4, 185–207.

——— (2011): "Asymptotic Theory for Nonparametric Regression with Spatial Data," *Journal of Econometrics*, 165, 5–19.

Rolin, J. (1992): "Some Useful Properties of the Dirichlet Process," *CORE Discussion Paper, Université Catholique de Louvain.*

Rothenberg, T. (1971): "Identification in Parametric Models," *Econometrica*, 39, 577–591.

Roy, A. (1952): "Safety First and the Holding of Assets," *Econometrica*, 20, 431–449.

Ruppert, D. (2002): "Selecting the Number of Knots for Penalized Splines," *Journal of Computational and Graphical Statistics*, 11, 735–757.

Ruppert, D. and R. Carroll (1997): "Penalized Regression Splines," *Technical Report.*

Ruppert, D. and P. Wand (1994): "Multivariate Locally Weighted Least Squares Regression," *The Annals of Statistics*, 22, 1346–1370.

Samuelson, P. (1938): "A Note on the Pure Theory of Consumer's Behaviour," *Economica*, 5, 61–71.

——— (1948): "Consumption Theory in Terms of Revealed Preference," *Economica*, 15, 243–253.

——— (1950): "The Problem of Integrability in Utility Theory," *Economica*, 17, 355–385.

Schwartz, L. (1966): *Théorie des Distributions*, Hermann, 1st ed.

Sethuraman, J. (1994): "A Constructive Definition of Dirichlet Priors," *Statistica Sinica*, 4, 639–650.

Skorokhod, A. (1956): "Limit Theorems for Stochastic Processes," *Theory of Probability and its Applications*, 1, 261–290.

Slutsky, E. (1915): "Sulla teoria del biancio del consumator," *Giornale degli Economisti*, 51, 1–26.

Stone, C. (1994): "The Use of Polynomial Splines and their Tensor Products in Multivariate Function Estimation," *The Annals of Statistics*, 22, 118–184.

Straf, M. (1972): "Weak Convergence of Stochastic Processes with Several Parameters," Berkeley.

Theil, H. and H. Neudecker (1958): "Substitution, Complementarity, and the Residual Variation around Engel Curves," *The Review of Economic Studies*, 25, 114–123.

Thurstone, L. (1927): "A Law of Comparative Judgement," *Psychological Review*, 34, 273–286.

Tibshirani, R. (1996): "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, 58, 267–288.

Tuvaandorj, P. (2020): "Regression Discontinuity Designs, White Noise Models, and Minimax," *Journal of Econometrics*, 218, 587–608.

Ullah, A. (1988): "Non-Parametric Estimation of Econometric Functionals," *The Canadian Journal of Economics*, 21, 625–658.

Van Rosen, D. (2018): *Bilinear Regression Analysis: An Introduction*, Springer.

Vanmarcke, E. (1983): *Random Fields: Analysis and Synthesis*, Cambridge: MIT Press.

Varian, H. (1982): "The Nonparametric Approach to Demand Analysis," *Econometrica*, 50, 845–973.

——— (1990): "Goodness-of-Fit in Optimizing Models," *The Journal of Econometrics*, 46, 125–140.

——— (1992): *Microeconomic Analysis*, New York: W. W. Norton & Company, Inc., 3rd ed.

Vinod, H. and A. Ullah (1987): "Flexible Production Function Estimation by Nonparametric Kernel Estimators," *University of Western Ontario*.

Walras, L. (1874): *Éléments D'économie Politique Pure*, Lausanne.

Watson, G. (1964): "Smooth Regression Analysis," *The Indian Journal of Statistics*, 26, 359–372.

Wikle, C. and N. Cressie (1999): "A Dimension-Reduced Approach to Space-Time Kalman Filtering," *Biometrika*, 86, 815–829.

Willey, J., N. Fettig, and M. Hale (2017): "The Extent of Trafficking in the Supplemental Nutrition Assistance Program: 2012-2014," *Prepared by WRMA for the U.S. Department of Agriculture, Food and Nutrition Service*.

Wishart, J. (2010): "Kink Estimation in Stochastic Regression with Dependent Errors and Predictors," *Electronic Journal of Statistics*, 4, 875–913.

——— (2011a): "Minimax Lower Bound for Kink Location Estimators in a Nonparametric Regression Model with Long Range Dependence," *Statistics and Probability Letters*, 81, 1871–1875.

——— (2011b): "Nonparametric Estimation of Change-Points in Derivatives," *Doctoral Thesis at the University of Sydney*.

Yu, K. and M. Jones (1998): "Local Linear Quantile Regression," *Journal of the American Statistical Association*, 93, 228–237.

Zinde-Walsh, V. (2018): "Nonparametric Functionals as Generalized Functions," *arXiv: 1303.1435v1*.